AD_____

Award Number: DAMD17-02-1-0346

TITLE: Scanning the Human Genome for Novel Therapeutic Targets for Breast Cancer

PRINCIPAL INVESTIGATOR: Gregory J. Hannon, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

REPORT DATE: April 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2005	Annual (1 Apr 2004 - 31 Mar 2005)

4. TITLE AND SUBTITLE
Scanning the Human Genome for Novel Therapeutic Targets for Breast Cancer

5. FUNDING NUMBERS
DAMD17-02-1-0346

6. AUTHOR(S)
Gregory J. Hannon, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

E-Mail: hannon@cshl.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)
No abstract provided.

14. SUBJECT TERMS
Cancer biology, genetics, synthetic lethality, apoptosis

15. NUMBER OF PAGES
97

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

The broad goal of this project is to develop genome-wide RNAi approaches in mammals and to apply these to the discovery of new therapeutic targets for cancer. Specifically, we have generated and continue to build a library of short hairpin RNA expression constructs (shRNA) that ultimately correspond to every gene in the human and mouse genomes. These will be are presently available as a public resource and used internally to screen for genes that are essential to the survival of breast cancer cells but which are dispensable for the survival of normal cells. A subset of these might prove suitable as therapeutic targets for breast cancer therapy. During the course of funding, two things have become clear. First, although they were not in place at the time of submitting this application, we have largely developed the technologies necessary to pursue the above goal. Second, funding in the Innovator award falls far short of that necessary to achieve the goal. Relevant to the last point, we have been able to leverage the Innovator award with several other funding sources to create a program, which is capable of meeting the proposed goal.

BODY

Progress toward developing the technology necessary for genome-wide RNAi in mammals (these were funded in part by the Innovator award and also by P01 from the NCI)

Studies on the mechanism of RNAi

All of the technology described above was built upon studies of the RNAi mechanism. While these studies are funded by an R01, they benefit from the Innovator award, and it is acknowledged as general support for the P.I. We have made progress relevant to this goal, understanding in more depth the biochemistry of the RNAi pathway and working to apply this knowledge to the improvement of RNAi as a tool. Progress over the last year relevant to this effort can be seen in the attached manuscripts (Appendicies 1-10).

A genome-wide RNAi library

Over the last several years, there have emerged two major methods for triggering RNAi in mammalian cells. These are transient silencing using siRNAs or stable or transient silencing using shRNAs. Both of these approaches have been validated in numerous publications. In considering how to construct a genome-wide RNA library for human cells, we examined both options. Our choice of the latter reflects several factors. First and foremost, shRNA expression constructs can be propagated and thus provide a limitless supply of material for public distribution. Second, many phenotypes, especially those relevant to breast cancer, require examination of cells over a long time frame. Third, shRNAs offer the flexibility to examine the consequences of silencing both

in vitro and in vivo. Generation of the library is proceeding as a phased project with funding coming in part from the Innovator award and in part from other public and commercial sources (NCI, Merck, Oncogene Sciences, Genetech). No funding mechanism has been permitted to place any restrictions on library distribution.

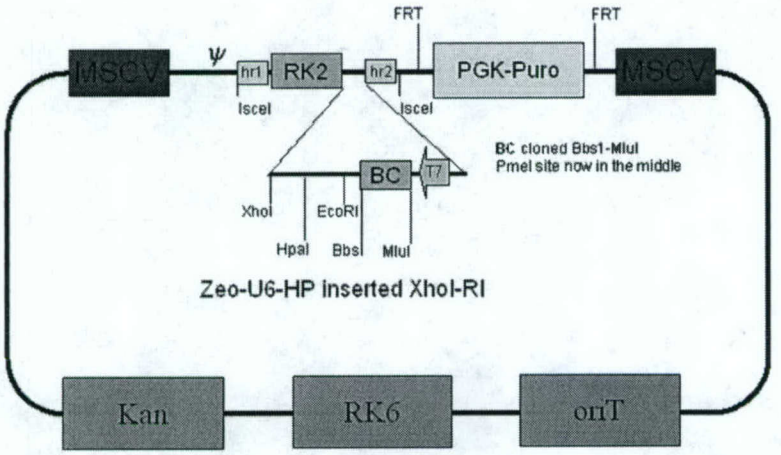
shRNA optimization

Our original series of vectors used a 29 base shRNA with a simple 4 base loop. Biochemical studies have revealed that natural miRNAs are processed into mature miRNAs through a two-step pathway that involves two specialized RNAseIII enzymes. We have found that by designing shRNAs as substrates for both components of the pathway that we can increase the amount of shRNA produced by our vectors in vivo by more than 10-fold. Additionally, detailed biochemical analysis of Dicer cleavage (see Appendix 6) has allowed us to apply informatic strategies for picking more efficient shRNA sequences. These innovations have been incorporated into our second generation library (see below). We have published a number of review articles as guides to the construction and use of shRNA reagents. These are attached as appendicies 2,4,5,8,9. Finally, we have provided examples of the use of shRNAs in vivo to study the role of the p53 tumor suppressor pathway (Appendix 3).

Vector construction and validation – vectors remain the same for both version 1 and version 2 libraries

In collaboration with Steve Elledge (Baylor), we constructed a flexible vector system for harboring the shRNA library. We have demonstrated that this vector can transfer shRNA inserts to a recipient plasmid by bacterial mating with ~100% efficiency. We have also validated transfers in multi-well formats suitable for moving subsets of or even the entire library. The original vector design had to be modified to remove loxP sites to avoid intellectual property restrictions placed upon us by Dupont. Persistant, although solvable problems, include the use of Zeocin and the inclusion of FRT sites. We did attempt negotiating with Salk and Invitrogen to overcome these barriers to distribution. However, terms remained unreasonable. However, this has been rendered moot by the speed with which we constructed the second generation library, which has been made available to the academic community prior to its publication. Nevertheless, we are now preparing a manuscript describing this new resource, and more than 500,000 of these plasmids have been distributed from Open Biosystems to academic investigators worldwide.

Figure 1. Vector system for library construction



For construction of the second generation library, we modified pSM1 (Figure1) slightly to incoporate flanking sequences from a naturally occurring human microRNA. We have also slightly rearranged the elements within the vector to increase its stability. The second generation vector, pSM2, is pictured below in Figure 2.

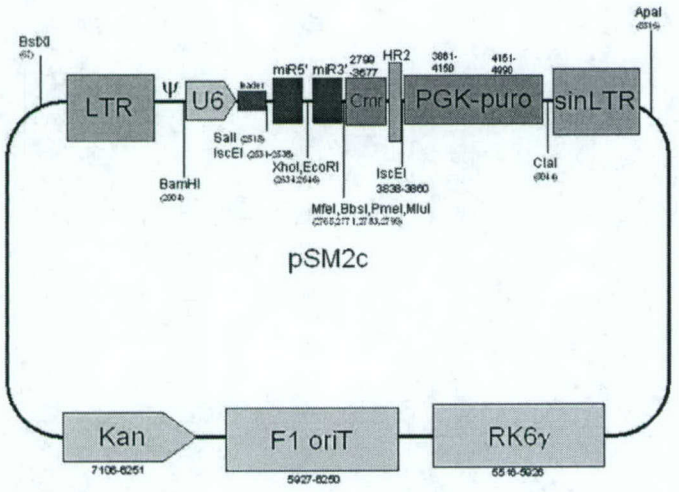


Figure 2. pSM2

shRNA library construction

Construction of the first generation library was completed and published last year. In the end, 9,780 genes were covered with 28,000 different shRNAs. Given our progress in understanding the RNAi pathway and our application of this knowledge to the generation of better shRNA tools, we have undertaken the construction of second-generation libraries that will cover all known and predicted genes in human, mouse and rat. Thus far, we have sequence verified second generation constructs that cover ~28,000 human genes with ~71,000 shRNAs. The ability to carry out construction on such a large scale required the development of novel oligonucleotide synthesis strategies, which take advantage of highly parallel piezo-inket technology for oligo production. A paper describing this process is attached as Appendix 7. We have also constructed and released about 51,000 mouse clones, covering 25,000 genes. In collaboration with Richard Gibbs and David Anderson, we have also initiated the production of a rat library.

Library Distribution

Our first and second generation libraries and vectors are now available to all academic investigators from Open Biosystems at reasonable cost.

Library Screening

We are currently carrying out library screens for activation of several reporters. This aspect of the grant has been delayed somewhat by the substantial effort required to construct and sequence verify the some 150,000 clones that now comprise our shRNA resource. However, we have completed one screen for new breast tumor suppressors in collaboration with Steve Elledge's lab. This identified a repressor protein involved in neuronal development as a gene that can contribute to epithelial carcinogenesis. This manuscript is attached as appendix 10. We have also developed methods for screening shRNA libraries by in situ transfection as described in Appendix 1.

KEY RESEARCH ACCOMPLISHMENTS

- Produced and publicly distributed a second-generation shRNA library consisting of 71,000 shRNAs to 28,300 human genes
- Produced and publicly distributed a second-generation shRNA library consisting of 51,500 shRNAs to 25,000 mouse genes
- validated new oligonucleotide synthesis methods for library construction

REPORTABLE OUTCOMES

- Manuscripts (8 attached)

Appendices :

1. Silva, J. M., Mizuno, H., Brady, A., Lucito, R. and Hannon, G. J. (2004) RNAi microarrays : High-throughput loss-of-function genetics in mammalian cells. Proc. Natl. Acad. Sci., 101:6548-6552.
2. Hemann, M. T., Zilfou, J. T., Zhao, Z., Burgess, D., Hannon, G. J. and Lowe, S. W. (2004) Suppression of tumorigenesis by the p53 target PUMA. Proc. Natl. Acad. Sci. 101:9333-9338.
3. Hannon, G. J. and Rossi, J. J. (2004) RNAi: unlocking the potential of the human genome. Nature 431:371-378.
4. Siolas, D., Lerner, C., Burchard, J., Ge, W., Paddison, P. J., Linsley, P., Hannon, G. J. and Cleary, M. Synthetic shRNAs as highly potent RNAi triggers. Nature Biotech. 23:227-231.
5. Cleary, M., Kilian, K., Wang, Y., Bradshaw, J., Ge, W., Cavet, G., Paddison, P. J., Chang, K., Sheth, N., Leproust, E., Coffey, E. M., Burchard, J., Linsley, P. and Hannon, G. J. (2004) Production of complex libraries of defined nucleic acid sequences using highly parallel in situ oligonucleotide synthesis. Nature Methods 1:241-248.
6. Paddison, P. J., Cleary, M., Silva, J. M., Chang, K., Sheth, N., Sachidanidan, R. and Hannon, G. J. (2004) Cloning of short hairpin RNAs for gene knockdown in mammalian cells. Nature Methods 1:163-167.
7. Silva, J., Chang, K., Hannon, G. J. and Rivas. F. V. (2004) RNA-interference-based functional genomics in mammalian cells : reverse genetics comes of age. Oncogene 23:8401-8409.
8. Westbrook, T. F., Martin, E. S., Schlabach, M. R., Leng, Y., Liang, A. C., Feng, B., Zhao, J. J., Roberts, T., Mandel, G., Hannon, G. J., DePinho, R., Chin, L. and Elledge, S. J. (2005) A genetic screen for candidate tumor suppressors identifies TGF-βRII and REST. Cell, in press.

- Second-generation human library ~2/3 complete
- Second-generation mouse library ~1/2 complete
- Second-generation rat library, in progress
- Design and public release of informatic tools for accessing the collection (codex.cshl.edu)

CONCLUSIONS

RNAi has emerged over the last two years as a powerful tool for experimental manipulation of gene expression and as a potential therapeutic strategy. We have made substantial progress toward validating the use of RNAi in mammals and have contributed key reagents to the scientific community.

RNA interference microarrays: High-throughput loss-of-function genetics in mammalian cells

Jose M. Silva, Hana Mizuno, Amy Brady, Robert Lucito, and Gregory J. Hannon*

Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

Edited by Stanley Fields, University of Washington, Seattle, WA, and approved March 3, 2004 (received for review January 9, 2004)

RNA interference (RNAi) is a biological process in which a double-stranded RNA directs the silencing of target genes in a sequence-specific manner. Exogenously delivered or endogenously encoded double-stranded RNAs can enter the RNAi pathway and guide the suppression of transgenes and cellular genes. This technique has emerged as a powerful tool for reverse genetic studies aimed toward the elucidation of gene function in numerous biological models. Two approaches, the use of small interfering RNAs and short hairpin RNAs (shRNAs), have been developed to permit the application of RNAi technology in mammalian cells. Here we describe the use of a shRNA-based live-cell microarray that allows simple, low-cost, high-throughput screening of phenotypes caused by the silencing of specific endogenous genes. This approach is a variation of "reverse transfection" in which mammalian cells are cultured on a microarray slide spotted with different shRNAs in a transfection carrier. Individual cell clusters become transfected with a defined shRNA that directs the inhibition of a particular gene of interest, potentially producing a specific phenotype. We have validated this approach by targeting genes involved in cytokinesis and proteasome-mediated proteolysis.

RNA interference (RNAi) has emerged as one of the standard techniques to study gene function in diverse experimental systems. Introduction of double-stranded RNA (dsRNA) into a cell decreases the level of the complementary mRNAs producing a knockdown of the corresponding protein. The current model of the RNAi mechanism proposes that the silencing "trigger" is processed by Dicer into small RNAs of 21–22 nucleotides in length. These become incorporated into an RNA-induced silencing complex with endonuclease activity (RISC), which, in turn, identifies and cleaves homologous mRNAs (1, 2).

Based on this approach, genomewide RNAi approaches have been used successfully for phenotype-based screens in *Caenorhabditis elegans* (3–5) and *Drosophila melanogaster* (6, 7). In part, these successes derive from the availability of convenient and inexpensive methods for producing and introducing dsRNA. For example, it has previously been shown that RNAi can be triggered by soaking *C. elegans* in a solution of dsRNA (8), or by feeding worms with *E. coli* expressing gene-specific dsRNAs (9). In *Drosophila* cells a soaking protocol is also available allowing an easy method of introducing dsRNA (10).

Unfortunately, similarly straightforward approaches for triggering silencing have not been described in mammals. Analysis of multiples genes requires a "gene by gene" method, in which individual transfections must be performed, making these studies expensive, tedious, and dependent on high-throughput robotic systems. Cell microarrays represent a novel alternative to classical approaches to phenotype-based assays in mammalian cells.

Cell microarrays were first described by Ziauddin and Sabatini (11), who demonstrated that cells grown on a glass substrate could take up DNA–lipid complexes that had been deposited on the slide before cells were plated. Cells essentially became transfected *in situ*, with defined spots of transfected cells localized over the printed DNAs. These studies demonstrated the use of conventional DNA constructs for creating phenotypes based on ectopic expression. Here we investigate the possibility of similarly using cell microarrays for loss-of-function genetics. This is accomplished by creating a microarray of living cells that have been transfected *in situ* with either small interfering RNAs (siRNAs) or with DNA constructs that direct the expression of short hairpin RNAs (shRNAs). These are effective at initiating a silencing response and in creating defined areas (spots) of cells in which suppression of a targeted gene generates an expected phenotype. Such arrays will find broad application to high-throughput low-cost phenotype-based screens in mammalian cells.

Materials and Methods

Microarray Printing and Reverse Transfection. Transfection mixes containing DNA reporter vectors (500 ng) plus shRNAs (1 μg) or siRNAs (200 ng) were printed onto glass slides by using a previously described "lipid method" (11) with some modifications. Briefly, nucleic acids were resuspended in 15 μl of DNA-condensation buffer (Buffer EC, Qiagen, Valencia, CA) with a final concentration of 0.4 M sucrose. After two incubation steps with the enhancer solution and the Effectene transfection reagent (Qiagen), a 1× volume of 0.2% gelatin was remixed with the solution to complete a transfection master mix of 45 μl. Ten microliters of this was aliquotted into a 384-well plate for printing, and the remaining 35 μl was stored at 4°C for later assays. Samples were printed onto Corning GAPS II slides with a PixSys 5500 Robotic Arrayer (Genomic Solutions, Ann Arbor, MI). Pins transferred the "lipid–DNA" solution to the slide while touching the surface of the slide for 500 ns. To ensure enough printed area to contain several hundred cells, we printed in close proximity nine spots forming a 3 × 3 square. After printing, the nine spots fuse together forming a bigger single dot with a diameter of ≈400–500 μm. To prevent contamination after printing, the slides were dried overnight at room temperature in a tissue culture hood.

To perform the reverse transfection, one spotted array was placed inside a 10-cm tissue culture dish and 15 ml of media containing cells at a concentration of 1×10^6 per milliliter was added into the dish taking care not to disturb the printed surface. Cells were incubated for 60 h without media change before analysis of the results (Fig. 1).

Reporter Assays. One hundred sixty dots containing a dual reporter vector expressing GFP/dsRed fluorescent proteins (gift of Alla Karpova, Cold Spring Harbor Laboratory) and individual shRNAs were printed. All shRNA were part of a library of U6 polymerase III promoter-driven hairpins (28). Four groups of experiments with 40 dots (each) were printed: the first group contained only dual reporter vector, the second group contained the reporter vector plus an shRNA or siRNA against firefly luciferase (Ff shRNA and Ff siRNA), the third group contained

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Ff, firefly luciferase; RNAi, RNA interference; shRNA, short hairpin RNA; siRNA, small interfering RNA.

*To whom correspondence should be addressed. E-mail: hannon@cshl.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0400165101

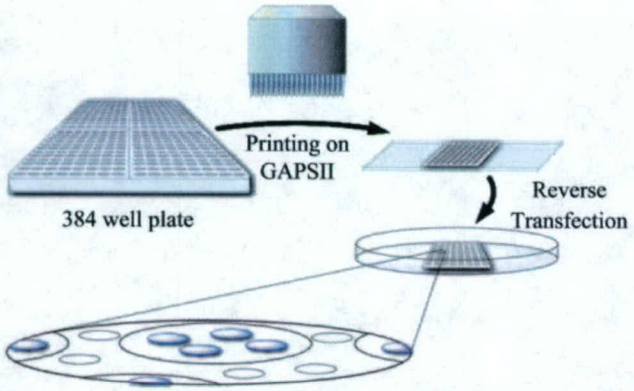


Fig. 1. Outline of the protocol used to perform reverse transfection on a glass slide. Transfection mixes containing gelatin were spotted on GAPS II glass slides with a PixSys 5500 Robotic Arrayer to form a 3 × 3 square. Dried slides were put inside a 10-cm tissue culture dish, and medium with cells was added to the dish. After incubation, groups of transfected cells could be detected inside the spotted transfection mixes.

the reporter vector plus a shRNA or a siRNA against GFP that has no effect in the expression level of the protein (GFP shRNA-1 and GFP siRNA-1), and the last group contained the reporter vector plus a shRNA that reduces by 90% the GFP signal when tested in culture plates (GFP shRNA-2 and GFP siRNA-2).

Several cell lines were tested for transfection, NIH 3T3, IMR90/E1A, HeLa, and HEK 293T. To test the stability of the printed array, we repeated the assay at different time points after printing, day 0, 1 week, 2 weeks, 4 weeks, and 2 months. For testing the stability of the transfection master mix, we stored the solution at 4°C and then printed new slides and assayed them at the time points described above.

Proteasome-Mediated Proteolysis Assays. Thirty shRNAs targeting different proteasome subunits were printed in triplicate. Every dot harbored an shRNA-expression vector, a plasmid expressing dsRed (dsRed N-1, Clontech), and a vector encoding a proteasome fluorescent reporter (ZsProSensor, Clontech). This reporter encodes a fusion protein that has been engineered to show varying levels of expression depending on the status of the proteasome pathway. Every transfection master mix contained 400 ng of dsRed vector, 100 ng of ZsProSensor, and 1 μg of shRNA plasmid. Twenty micrograms of total protein lysates was used for Western blot analysis. Rabbit anti-PSMC-6 subunit of the proteasome (Affinity, Biomol, Plymouth Meeting, PA), rabbit anti-ubiquitin (StressGen Biotechnologies, Victoria, Canada), and mouse anti-β-actin (United States Biological, Swampscott, MA) antibodies were also used in these studies.

Cytokinesis Defect Assays. Eight shRNAs targeting the motor protein Eg5 were printed (10 replicas each) together with a vector encoding an α-tubulin-GFP fusion protein (GFP-tubulin, Clontech). Every transfection mix contained 1 μg of Eg5 hairpin and 500 ng of the fluorescent fusion protein expression plasmids.

For immunofluorescence studies, we printed a replica slide where the transfection mix contained 500 ng of the dsRed reporter instead of the GFP-tubulin. Cells were stained by using standard methods with small variations. After incubation of the slides for 60 h the cells were fixed with paraformaldehyde for 10 min, washed very gently, and permeabilized with 1% Triton X-100 in PBS for 15 min on ice. Only one 15-min wash was performed to avoid washing away the cells. Mouse anti-α-tubulin (Sigma) was used in this study. Hoechst dye was included in the last wash to visualize the chromosomes.

Ninety-Six-Well Plate Analyses. All RNAi microarray results were validated by using cells transfected in 96-well tissue culture plates. Cells were transfected with LT-1 (Mirus, Madison, WI) according to the manufacturer's instructions at 50–70% confluence. The plasmids containing appropriate constructs were cotransfected, keeping the same ratios used in the arrayed slides but with a total mass of 100 ng of DNA for each transfected well. Again, results were analyzed after 60 h of incubation.

Results

Targeting Reporter Genes *in Situ* by Using siRNAs. Given previous successes in ectopically expressing genes by reverse transfection (11), we hoped that similar approaches could be coupled with the use of RNAi to produce knockdown phenotypes. Therefore, we began by testing the ability of siRNAs to be deposited on a microarray as lipid–RNA complexes and to cause sequence-specific silencing in cells grown on the arrays. We began by testing the ability of siRNAs to silence a co-delivered, ectopic marker. For convenience, we used GFP to enable both transfection and silencing to be scored by visual inspection. Because GFP was the siRNA target, we also included a plasmid that directs the expression of a second fluorescent protein (dsRed) to allow us to verify that siRNAs specifically silenced GFP expression rather than simply interfered with transfection.

Several siRNAs homologous to the GFP coding sequence were mixed with plasmids encoding GFP and dsRed. Nucleic acids were combined with a variety of lipid reagents [LT-1 or TKO (Mirus) and Effectene (Qiagen)], and the lipid–nucleic acid complexes were spotted onto glass slides. We found that Qiagen reagent performed best, giving optimal transfection of both DNA and RNA in this experiment. GFP siRNAs that had been previously identified as an efficient suppressor of the expression of the fluorescent protein in standard transfections (GFP siRNA-2) also showed potency on the spotted array. Ineffective siRNAs (GFP siRNA-1) or unrelated siRNAs (Ff siRNA) did not produce any effect on GFP expression levels either in standard transfections or on microarray slides (Fig. 2*A*).

Targeting Reporter Genes *in Situ* by Using shRNAs. The aforementioned studies demonstrated that an RNAi response could be initiated by *in situ* transfection of siRNAs on glass-slide microarrays. As stated above, RNAi can also be initiated in mammalian cells by transfection of DNA expression constructs that direct the synthesis of shRNA sequences. We therefore tested the possibility that a slide printed with a transfection mix containing an shRNA-expression construct that targets a specific mRNA could initiate a specific silencing response in *in situ* transfected cell clusters. As a first step, we again tested our ability to down-regulate the expression of GFP. Spots were printed as described above. In every cell cluster, the GFP level reported specific RNAi-mediated suppression and the dsRed level again acted as a transfection control.

As expected, none of the spots containing the reporter vector alone, the reporter plus a control shRNA, targeting firefly luciferase, or the reporter plus GFP shRNA-1, an ineffective shRNA, showed any reduction in the expression of GFP. In contrast, all spots harboring the GFP shRNA-2, an active shRNA, presented a strong attenuation of the GFP signal while maintaining unaltered levels of dsRed protein (Fig. 2*B*). It is worth noting that all suppressed samples showed a similar degree of attenuation of GFP and a very similar percentage of transfected cells, revealing a high degree of consistency from spot to spot.

We next asked whether the *in situ* RNAi procedure could be applied broadly by examining the response of a number of different cell lines in an arrayed format. These included transformed and nontransformed cells lines, fibroblasts and epithelial cells, and lines from mouse and human. Of those tested, HEK

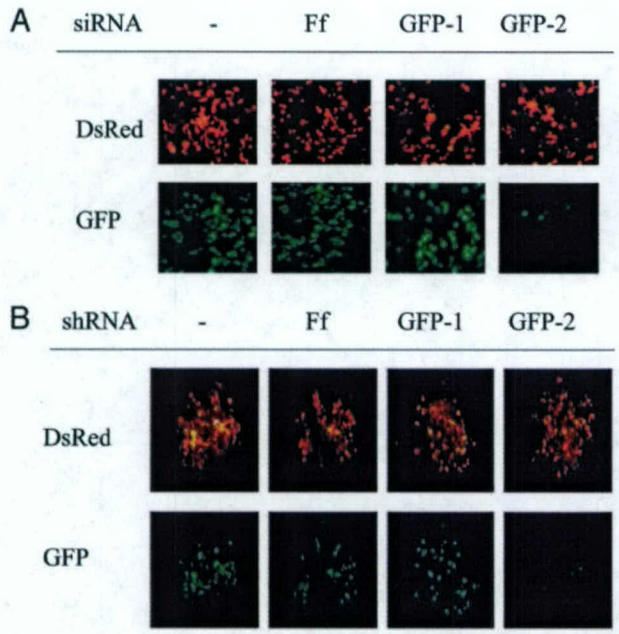


Fig. 2. The GFP reporter is specifically suppressed by RNAi in 293T cells incubated for 48 h on a printed glass slide. (*A*) Printed spots contained a vector expressing GFP and dsRed reporters and individual siRNAs targeting different sequences (as indicated): no siRNA (—), Ff, siRNA-1 (GFP-1), and siRNA-2 (GFP-2). (*B*) Printed spots contained a vector expressing GFP and dsRed reporters and individual shRNA expression plasmids targeting different sequences (as indicated): no hairpin (-), Ff, hairpin 1 (GFP-1), and hairpin 2 (GFP-2). Control spots harbored siRNA or an shRNA-expression vector against Ff, GFP-1, or no added RNAi inducer (mock).

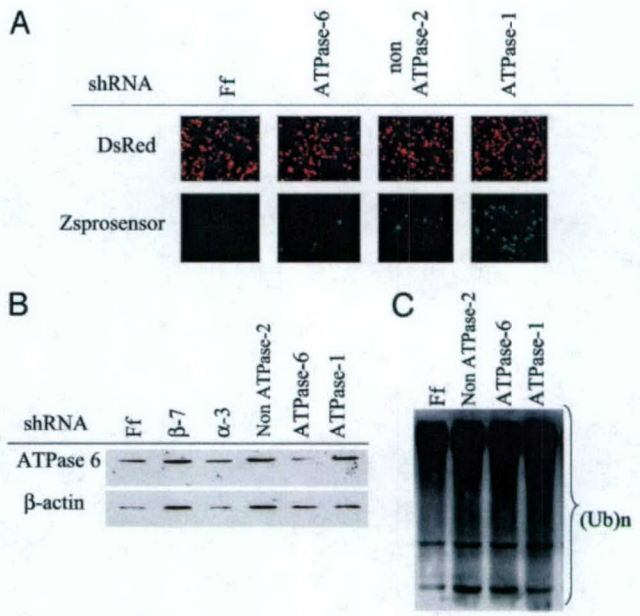


Fig. 3. (*A*) Different levels of ZsGreen protein accumulation detected in the RNAi microarray. (*B*) HEK 293T were incubated with individual hairpins targeting the expression of different proteasome subunits. A Western blot shows specific inhibition of ATPase-6 shRNA. Firefly shRNA was used as control. (*C*) Anti-ubiquitin Western blot showing accumulation of polyubiquitinated proteins in treated cells.

293T cells showed the highest efficiency. IMR90/E1A, NIH 3T3, and HeLa cells showed lower efficiency (20–50%) than 293T (data not shown). Our results indicate that by varying the lipid and nucleic acid content of the spot, many different cell lines can be used. However, for convenience, we performed the remainder of our assays with HEK 293T.

We also examined the stability of slides spotted with lipid/shRNA mixtures and with lipid/siRNA mixtures. We printed replica slides and compared both transfection efficiency and silencing efficiency at various time points. We did not observe any reduction in the performance of the shRNA arrays after 1 week, 1 month, or up to 2 months of storage at 4°C. In contrast, siRNA printed slides showed more variability, and unclear results were obtained after 2 weeks of storage.

An *in Situ* Assay for Defects in Proteosome Function. The need to locate transfected cells to score phenotypes on the array led to the consideration of phenotypic assays that depended on the expression of a fluorescent, exogenous reporter gene. The basic idea, a variation of the validation experiments described above, was to include in the transfection mix a plasmid harboring a fluorescent reporter that is differentially expressed (at higher or lower level) in cells in which RNAi has been used to alter the function of a specific pathway. To test this approach, we focused on protein half-life as a determinant of steady-state expression levels. A variety of motifs have been found to confer a short lifetime on cellular proteins. So-called PEST sequences are rich in the amino acids Pro (P), Asp, Glu (E), Ser (S), and Thr (T), which occur in internal positions in the sequence. Proteins containing PEST sequence elements are rapidly targeted to the 26S proteasome for degradation (12).

The mouse ornithine decarboxylase (MODC) has an extremely short half-life (13). MODC degradation is mediated by an internal domain, called MODC-d410, that contains several PEST motifs. This functional motif is transferable, decreasing greatly the stability of proteins to which it is fused (14, 15). This property has been exploited to create commercially available reporter systems in which MODC-d410, appended to a fluorescent protein, creates a fusion that can indicate the integrity of the proteosome in living cells.

We chose to analyze the expression level of a commercially available reporter consisting of a green fluorescent protein (ZsGreen) tagged on the carboxyl-terminus with the MODC-d410 domain (ZsProSensor, Clontech). We recently screened a library of 7,000 individual shRNAs for the ability to antagonize proteosome function in a 96-well plate format (28). Roughly one-quarter of primary positives targeted proteosome subunits. Putative positives from that screen were used on microarrays to validate the *in situ* approach and to compare the sensitivity of array-based assays to those carried out in 96-well plates. Based on these previous studies, thirty different hairpins targeting proteasome subunits were deposited, each in individual spots, together with the destabilized green reporter and a dsRed vector. As in previously described experiments, dsRed served as transfection control.

After 24 h of incubation we observed higher levels of ZsGreen protein in several dots containing proteasome shRNAs compared to control shRNAs, whereas no changes were observed in the intensity of the red fluorescent protein. During the following 48 h the green fluorescent signal gradually increased in these positive spots, achieving maximum intensity after 60–72 h. We identified clear differences in the accumulation level of the reporter in spots containing proteasome shRNAs compared with nonproteasome hairpins. Interestingly, the most intense signals were revealed from spots containing shRNAs that target subunits of the 19S base (Fig. 3*A*).

To confirm that the increased expression of the reporter was associated with alterations in proteasome function, we verified that cells transfected with a hairpin against the ATPase subunit 6 indeed showed a specific reduction of the targeted protein. As expected, Western blot analysis of transiently transfected 293T

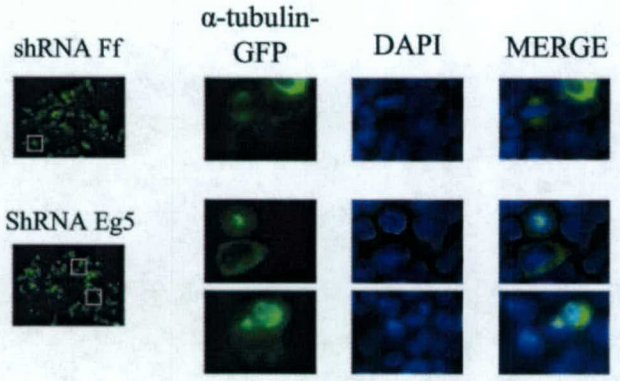


Fig. 4. Cytokinesis defects induced by Eg5 shRNAs. (*Upper*) Sample of normal mitotic metaphase detected in a dot containing firefly (control) shRNA. (*Lower*) Two typical "rosette" phenotype found in a printed spot containing the shRNA-7 against the motor protein Eg5.

cells with the ATPase-6 shRNA-1 showed a specific reduction in the level of the targeted protein (Fig. 3*B*).

It has been shown previously that in cells in which the normal function of the proteasome is blocked, there is an accumulation of polyubiquitinated proteins (16, 17). To validate further the antagonism of proteosome function by RNAi, we examined the level of polyubiquitinated proteins. Indeed, analysis of bulk ubiquitinated proteins by Western blotting with a ubiquitin antibody revealed an increase in these normally unstable species in cells treated with proteasome shRNAs compared with controls (Fig. 3*C*).

A Screen for Alterations in Cell Cycle Control: Cytokinesis Defects. The aforementioned screen was amenable to the use of microarray scanners to score positives by their ratio of fluorescent signals. To examine the suitability of this approach for other types of screens, we performed a live-cell microarray assay to identify cytokinesis defects induced by RNAi. As a reconstruction experiment, we knocked down the mitotic motor protein Eg5, because cytokinesis defects in cells where Eg5 function is inhibited are well established (18, 19).

The mitotic spindle, which consists of a dynamic array of microtubules and associated proteins, is responsible for segregation of chromosomes during mitosis. Studies using immunodepletion (20, 21) of the motor kinesin Eg5 or with the specific inhibitor monastrol (18) have revealed that inhibition of the normal function of this protein causes a defect in spindle formation. Initially, a defect in centrosome separation causes the assembly of monopolar spindles, and eventually, aberrant structures form that look like "rosettes" of microtubules with DNA at the periphery.

Based on the expectation of this typical morphology, we printed DNA–lipid complexes containing individual shRNAs targeting Eg5 on microarray slides. Transfection mixes also contained a plasmid encoding an α-tubulin GFP fusion protein. In this array, the GFP fusion protein identifies the cells that have been transfected and allows visualization of microtubules. Additionally, when the microarray was scored, Hoechst 33342 was added to the media to allow visualization of chromosomes. The analysis of the array revealed two hairpins that produced cells displaying the characteristic "rosette" pattern (Fig. 4). We did not observe this phenotype in any of the spots containing control shRNAs. This phenotype was also observed when shRNAs targeting Eg5 were tested in a 96-well plate format and was similar to the phenotype obtained after Eg5 inhibition by monastrol (data not shown).

The use of vectors harboring reporters or fusion proteins is a very convenient approach to identify abnormal phenotypes by *in situ* transfection. However, appropriate reagents will not be easily available for all interesting phenotypes. For this reason, we asked whether it was possible to identify Eg5 suppression phenotypes by the conventional technique of immunofluorescence (IF). A replica of the Eg5 glass slide, in which the α-tubulin fluorescence reporter was replaced by dsRed, was stained with a standard IF protocol. We find that we could easily identify cytokinesis defects produced by the same shRNAs identified with the α-tubulin GFP protein by using this methodology (data not shown).

Discussion

Genomewide analyses of loss-of-function phenotypes in mammals, similar to classic genetic studies in yeast, were very difficult, if not impossible, only a few years ago. However, over the last several years RNAi has emerged as a powerful approach for manipulating gene expression in mammalian cells, opening the door to the execution of such screens. High-throughput RNAi analyses have previously been used to study gene function in *D. melanogaster* and *C. elegans*. For example, essential genes (22), G protein-coupled receptors (GPCRs) (23), fat regulatory genes (24), and genes that regulate lifespan (25) have been functionally analyzed by this approach. Unfortunately, similar studies in mammals still represent a technological challenge. In part, limitations occur because of the cost of the RNA species themselves and the cost of introducing these species (siRNA and shRNA) into mammalian cells. Additionally, the use of a large-scale analysis in any system is limited by the screening methodology.

Here we validate the use of a cost-effective high-throughput procedure for RNAi-based screens in mammalian cells. This procedure is based on methodologies developed by Sabatini and colleagues (11) for creating high copy suppression phenotypes by "reverse transfection." This allows for the cost-effective use of materials, both the nucleic acids themselves and tissue culture reagents. We estimate that between 100 and 500 reverse transfections can be done with the materials required for a single transfection in a well of a 96-well plate. Additionally, thousands of samples can be printed in parallel on a glass-slide microarray, reducing the time and cost associated with maintaining cultures and analyzing phenotypic outputs. Finally, as previously described, printed slides can be stored for several months without losing potency (11).

By comparing results obtained by initiating RNAi *in situ* on microarrays to screens conducted in 96-well plates, we find that the arrays compare favorably to standard methods for both sensitivity and specificity. In agreement with our data, two recent papers have reported that RNAi could be initiated on printed slides to inhibit the expression of a co-delivered marker gene (26, 27). Our study extends these results by showing that a similar procedure can be used to silence endogenous genes to create RNAi-induced phenotypes relevant to two independent biological pathways.

Firstly, we designed an assay to study proteosome-mediated degradation. As predicted, when subunits of the 26S proteasome were targeted by shRNAs, we could clearly identify accumulation of an engineered protein that is normally degraded by the proteasome pathway. We showed that proteolytic defects were due to specific suppression of the targeted 26S subunits and demonstrated that cellular levels of the natural proteasome substrates (ubiquitinated proteins) were affected in the same way as the fluorescent reporter. In a second study, we analyzed the effect produced by loss of kinesin Eg5 as model for cytokinesis defects. After targeting with Eg5 shRNAs, we could reproduce the expected aberrant spindle morphology in a printed slide format, while no changes were observed in control spots.

Here, we demonstrate the feasibility of using printed arrays of siRNAs and shRNAs for highly parallel phenotype analysis in living cells. This approach is flexible and provides low-cost alternatives to similar screens carried out in multiwell plate formats. As large libraries of shRNAs become widely available

(28, 29), the techniques described herein will become a powerful approach to genetic analysis in mammalian cells.

We thank Jim Duffy for help with preparation of the figures and Michelle Carmell, Ahmet Denli, Liz Murchinson, Lin He, and William M. Keyes for critical reading of the manuscript. J.M.S. is supported by a postdoctoral fellowship from the U.S. Army Prostate Cancer Research Program, and G.J.H. is supported by an Innovator Award from the U.S. Army Breast Cancer Research Program. This work was supported in part by grants from the National Institutes of Health.

1. Zamore, P. D. (2001) *Nat. Struct. Biol.* **8,** 746–750.
2. Hannon, G. (2002) *Nature* **418,** 244–251.
3. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. (2001) *Curr. Biol.* **11,** 171–176.
4. Ashrafi, K., Chang, F. Y., Watts, J. L., Fraser, A. G., Kamath, R. S., Ahringer, J. & Ruvkun, G. (2003) *Nature* **421,** 268–272.
5. Pothof, J., van Haaften, G., Thijssen, K., Kamath, R. S., Fraser, A. G., Ahringer, J., Plasterk, R. H. & Tijsterman, M. (2003) *Genes Dev.* **17,** 443–448.
6. Somma, M. P., Fasulo, B., Cenci, G., Cundari, E. & Gatti, M. (2002) *Mol. Biol. Cell.* **13,** 2448–2460.
7. Kiger, A., Baum, B., Jones, S., Jones, M., Coulson, A., Echeverri, C. & Perrimon, N. (2003) *J. Biol.* **2,** 27.
8. Tabara, H., Grishok, A. M. &. Mello, C. C. (1998) *Science* **282,** 430–431.
9. Timmons, L. & Fire, A. (1998) *Nature* **395,** 854.
10. Clemens, J. C., Worby, C. A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B. A. & Dixon, J. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6499–6503.
11. Ziauddin, J. & Sabatini, D. M. (2001) *Nature* **411,** 107–110.
12. Rechsteiner, M. & Rogers, S. W. (1996) *Trends Biochem. Sci.* **21,** 267–271.
13. Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S., Igarashi, K., Tamura, T., Tanaka, K. & Ichihara, A. (1992) *Nature* **360,** 97–99.
14. Olmo, M. T., Rodriguez-Agudo, D., Medina, M. A. & Sanchez-Jimenez, F. (1999) *Biochem. Biophys. Res. Commun.* **257,** 269–272.
15. Verma, R. & Deshaies, R. J. (2000) *Cell* **101,** 341–344.
16. Wojcik, C. & DeMartino, G. N. (2002) *J. Biol. Chem.* **277,** 6188–6197.
17. Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. & Huber, R. (1999) *Annu. Rev. Biophys. Biomol. Struct.* **28,** 295–317.
18. Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L. & Mitchison, T. J. (1999) *Science* **286,** 971–974.
19. Goshima, G. & Vale, R. D. (2003) *Cell Biol.* **162,** 1003–1016.
20. Sawin, K. E., LeGuellec, K., Philippe, M. & Mitchison, T. J. (1992) *Nature* **359,** 540–543.
21. Blangy, A., Lane, H. A., d'Herin, P., Harper, M., Kress, M. & Nigg, E. A. (1995) *Cell* **83,** 1159–1169.
22. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. (2001) *Curr. Biol.* **11,** 171–176.
23. Keating, C. D., Kriek, N., Daniels, M., Ashcroft, N. R., Hopper, N. A., Siney, E. J., Holden-Dye, L. & Burke, J. F. (2003) *Curr. Biol.* **13,** 1715–1720.
24. Ashrafi, K., Chang, F. Y., Watts, J. L., Fraser, A. G., Kamath, R. S., Ahringer, J. & Ruvkun, G. (2003) *Nature* **421,** 268–272.
25. Lee, S. S., Lee, R. Y., Fraser, A. G., Kamath, R. S., Ahringer, J. & Ruvkun, G. (2003) *Nat. Genet.* **33,** 40–48.
26. Mousses, S., Caplen. N. J., Cornelison, R., Weaver, D., Basik, M., Hautaniemi, S., Elkahloun, A. G., Lotufo, R. A., Choudary, A., Dougherty, E. R., *et al.* (2003) *Genome Res.* **13,** 2341–2347.
27. Kumar, R., Conklin, D. S. & Mittal, V. (2003) *Genome Res.* **13,** 2333–2340.
28. Paddison, P. J., Silva, J. M., Conklin, D. S., Schlabach, M., Li, M., Aruleba, S. Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., *et al.* (2004) *Nature* **428,** 427–431.
29. Berns, K., Hijmans, E. M., Mullenders, J., Brummelkamp, T. R., Velds, A., Heimerikx, M., Kerkhoven, R. M., Madiredjo, M., Nijkamp, W., Weigelt, B., *et al.* (2004) *Nature* **428,** 431–437.

Suppression of tumorigenesis by the p53 target PUMA

Michael T. Hemann*†, Jack T. Zilfou*†, Zhen Zhao‡, Darren J. Burgess*, Gregory J. Hannon*, and Scott W. Lowe*§

*Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724; and ‡Genetics Program, Stony Brook University, Stony Brook, NY 11794

Communicated by Bruce W. Stillman, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, May 10, 2004 (received for review April 4, 2004)

The p53 tumor suppressor regulates diverse antiproliferative processes such that cells acquiring *p53* mutations have impaired cell-cycle checkpoints, senescence, apoptosis, and genomic stability. Here, we use stable RNA interference to examine the role of PUMA, a p53 target gene and proapoptotic member of the Bcl2 family, in p53-mediated tumor suppression. *PUMA* short hairpin RNAs (shRNAs) efficiently suppressed *PUMA* expression and p53-dependent apoptosis but did not impair nonapoptotic functions of p53. Like *p53* shRNAs, *PUMA* shRNAs promoted oncogenic transformation of primary murine fibroblasts by the *E1A*/*ras* oncogene combination and dramatically accelerated *myc*-induced lymphomagenesis without disrupting p53-dependent cell-cycle arrest. However, the ability of PUMA to execute p53 tumor suppressor functions was variable because, in contrast to *p53* shRNAs, *PUMA* shRNAs were unable to cooperate with oncogenic *ras* in transformation. These results demonstrate that the p53 effector functions involved in tumor suppression are context dependent and, in some settings, depend heavily on the expression of a single proapoptotic effector. Additionally, they demonstrate the utility of RNA interference for evaluating putative tumor suppressor genes *in vivo*.

The p53 tumor suppressor is a transcription factor that controls diverse cellular processes such as DNA repair, cell-cycle checkpoints, senescence, apoptosis, and angiogenesis (1). In principle, disruption of each of these activities alone or in combination could explain the potent impact of *p53* mutations on tumorigenesis. Attempts to identify relevant p53 activities have used mouse models or cells derived from these animals to determine whether disruption of individual p53 effectors can mimic *p53* loss during tumorigenesis. To date, the biological consequences of inactivating these effectors have not been as severe as those obtained by inactivating *p53* itself. For example, disruption of either *bax* (an apoptotic regulator) or *p21* (a proliferation inhibitor) does not recapitulate *p53* loss in promoting transformation or tumorigenesis (2–4). Although these observations suggest that disruption of multiple p53 functions is required to support tumorigenesis, it is also clear that p53 coordinates each activity through multiple effectors whose single inactivation is not sufficient to completely disable each activity (5).

Other studies have taken a more global approach to specifically target p53 activities, albeit with apparently contradictory results. For example, during *myc*-induced B cell lymphomagenesis, coexpression of *bcl2*, which completely disables apoptosis downstream of p53, mimics *p53* mutations in producing aggressive malignancies that retain p53-dependent cell-cycle checkpoints (6). Although these results suggest that apoptosis is the primary p53 activity responsible for tumor suppression in this model, p53 deficiency, but not expression of *bcl2*, is efficient at promoting T cell lymphomagenesis (7–9). Furthermore, mice harboring p53 point mutants that are incapable of transactivating p53 proapoptotic targets fail to develop the T cell lymphomas characteristic of p53-null mice (7, 10). In the latter setting, genomic instability, and not apoptosis, was proposed to explain the advantage of *p53* mutations during tumorigenesis.

PUMA (p53 up-regulated modulator of apoptosis) is a "BH3-only" member of the Bcl2 family that was initially identified from differential gene expression studies as a p53 target gene and a potent inducer of apoptosis (11, 12). PUMA acts by modulating Bax activity to facilitate cytochrome *c* release from the mitochondria, thereby triggering the apoptotic cascade (11). *PUMA*-deficient colon carcinoma cells and MEFs derived from *PUMA*-deficient mice are resistant to several apoptotic stimuli, including those acting through p53 (13, 14). In fact, the similar phenotypes of $PUMA^{-/-}$ and $p53^{-/-}$ cells suggest that PUMA is an essential p53 effector during apoptosis under some conditions. However, $PUMA^{-/-}$ mice are not overtly tumor-prone, again suggesting that simultaneous inactivation of multiple p53 effector functions is critical for tumorigenesis (15).

We have proposed that not all p53 effector functions contribute to tumor suppression, and, instead, that loss of specific p53 activities can play crucial roles that are context-dependent (6, 16). For example, whereas apoptosis appears to be the primary p53 activity limiting *myc*-induced lymphomagenesis, both apoptosis and cellular senescence contribute to chemotherapy responsiveness (6, 17). Because PUMA can be a specific and essential mediator of p53-dependent apoptosis, the impact of *PUMA* loss on malignant phenotypes should reveal the relative contribution of apoptosis to p53-mediated tumor suppression in different contexts. Therefore, we used stable RNA interference (RNAi) to acutely suppress *PUMA* expression in settings where p53 has established tumor-suppressor activity. In contrast to studies with traditional knockout mice, studies with RNAi exploit short hairpin RNAs (shRNAs) to acutely and stably suppress gene expression, providing an extremely rapid approach to study loss-of-function effects *in vitro* and *in vivo* (16, 18). Also, RNAi can produce hypomorphic expression states that may more closely mimic basal expression states or the activity of a mutated gene than a null allele. Here, we use this approach to identify PUMA as a potential tumor suppressor and highlight important features of the p53 tumor-suppressor network.

Materials and Methods

Cells and Gene Transfer. Primary murine embryonic fibroblasts were derived from WT and $p53^{-/-}$ day 13.5 embryos and maintained as described in refs. 19 and 20. Retroviruses encoding shRNAs expressed from the U6 promoter were generated by PCR with a pGEM U6 promoter template (16). The shRNA sequences encoded inverted repeats of 29 nt separated by an 8-nt spacer. The inverted repeats corresponded to nucleotides 772–802 (shPU-2) or 500–528 (shPU-3) of the mouse *PUMA* cDNA (NM_133234) and had >3-nt differences compared with any other murine genes as determined by BLAST. The shp53 sequence used was the same as the "p53.2" published in ref. 16. The resulting PCR products were cloned directly into the *Hpa*I site of the murine stem cell virus (MSCV) phosphoglycerate kinase (PGK)-Puro-internal ribosome entry site (IRES)-GFP vector

Abbreviations: HSC, hematopoietic stem cell; IRES, internal ribosome entry site; MEF, mouse embryo fibroblast; MSCV, murine stem cell virus; PGK, phosphoglycerate kinase; RNAi, RNA interference; shRNA, short hairpin RNA.

†M.T.H. and J.T.Z. contributed equally to this work.

§To whom correspondence should be addressed. E-mail: lowe@cshl.edu.

GENETICS

(16) or, for the *p53* shRNA vector used *in vitro*, pQCXIX PGK-Puro-IRES-GFP (modified from Clontech's pQCXIX). The exact primer sequences and cloning strategies are available from the authors upon request. All animal protocols were approved by Cold Spring Harbor Laboratory in accordance with National Institutes of Health guidelines.

Retroviral-mediated gene transfer was performed by using Phoenix packaging cells (G. Nolan, Stanford University, Stanford, CA) as described in ref. 20. Oncogenic *ras* (*H-RasV12*) (20) and *E1A* were expressed by using WZL-Hygro-based retroviral vectors (Ariad Pharmaceuticals, Cambridge, MA). *E1A/H-RasV12* was expressed by using a modified pBabe H-RasV12 retroviral vector (20). Infected cell populations were selected by culture in puromycin (2 μg/ml, 3 days) or hygromycin (100 μg/ml, 3 days) to eliminate uninfected cells.

PUMA Suppression and Functional Assays. Cells were extracted in RIPA buffer (50 mM Tris, pH 7.4/150 mM NaCl/1% Triton X-100, 0.1% SDS/1% sodium deoxycholate) supplemented with Complete Mini protease inhibitors (Roche Diagnostics). We assessed protein expression by immunoblotting as described in ref. 21, with primary antibodies directed against PUMA (1:300) from ProSci (San Diego), p53 (1:500) from NovoCastra (Newcastle, U.K.), or α-tubulin (1:4,000) from Sigma. Mouse embryo fibroblasts (MEFs) used in *E1A/ras* death assays were plated into 12-well plates (10^5 cells per well) in medium containing either 10% or 0.1% FBS and incubated for $\approx$36 h. Cell viability was analyzed by trypan blue exclusion, and at least 200 cells were scored for each sample.

For BrdUrd incorporation assays, 1×10^5 $p53^{-/-}$ MEFs or WT MEFs infected with MSCV control, shPU-3, or shp53 viruses were plated in triplicate on sterile coverslips in six-well dishes. Twelve hours after plating, MEFs were exposed to 6-Gy γ-radiation. Fourteen hours later, cells were pulsed with BrdUrd for 4 h. BrdUrd incorporation and colony formation assays were performed as described in ref. 16. For lymphoma DNA damage checkpoint analysis, tumor-bearing mice were either irradiated at 6 Gy or left untreated. Thirty-six hours later, lymphoma cells were harvested from irradiated and control mice, fixed in 70% ethanol, and stained with propidium iodide for cell-cycle analysis (6).

Tumorigenicity Assays. For tumorigenicity assays, 1×10^6 cells per 0.25 ml of PBS were injected s.c. into NIH Swiss athymic nude mice (Taconic Farms) and monitored as described in ref. 22. MEFs from two different embryo preparations were infected with retroviruses containing shRNAs targeting *PUMA* or *p53* in the presence of MSCV vector, *ras*, or *E1A/ras*. Isolation, infection, and transplantation of hematopoietic stem cells derived from embryonic day 15 WT or *Eμ-myc* was performed as described in ref. 6. Reconstituted animals were monitored for illness by lymph node palpation, by monitoring overall morbidity, and, in some cases, by whole-body fluorescence imaging (8). Overall survival was defined as the time from stem cell reconstitution until the animal reached a terminal stage and was killed. In all cases, terminal animals harbored large tumor burdens. Statistical analysis was performed with a one-way ANOVA test by using PRISM (Version 3.0, GraphPad, San Diego).

Immunophenotyping was performed on three shPUMA and three shp53-induced lymphomas. Briefly, 1×10^6 freshly harvested lymphoma cells were washed two times in PBS with 2% FBS. Cells were then incubated for 1 h in 200 μl of PBS with 10% FBS containing 1:100 dilutions of phycoerythrin-conjugated B220, Thy1, and IgM (Pharmingen). After incubation with surface antibodies, cells were washed two times in PBS with 2% FBS. Flow cytometry analysis was performed on a Becton Dickinson LSRII cell analyzer equipped with FACSVANTAGE DIVA software.

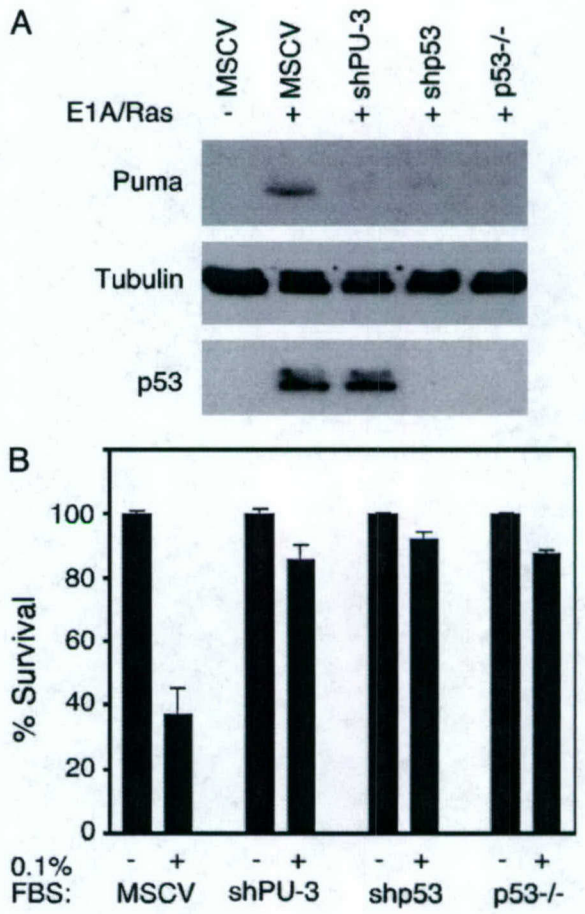


Fig. 1. Analysis of *PUMA* shRNA function *in vitro*. (*A*) Western blot analysis showing p53 and PUMA levels in *E1A/ras* MEFs stably expressing an MSCV control vector or *PUMA* and *p53* shRNAs. Tubulin is shown as a loading control. (*B*) *E1A/ras* MEFs transduced with an MSCV control vector or *p53* and *PUMA* shRNAs were incubated in high/low serum for 36 h after which viability was assessed by trypan blue exclusion.

Results

***PUMA* shRNAs Can Suppress PUMA Levels and Activity.** MEFs expressing *E1A* or *myc* become sensitized to apoptosis after DNA damage or serum depletion (5). In contrast, oncogene-expressing MEFs derived from both $p53^{-/-}$ and $PUMA^{-/-}$ mice are resistant to apoptosis (14–15, 23), suggesting that PUMA is essential for p53-mediated cell death in this context. To determine whether suppression of *PUMA* by using RNAi could be effective, we generated *PUMA* shRNAs against distinct sequences in the *PUMA* gene (designated shPU-2 and shPU-3), cloned them into a retroviral expression vector, and tested their activity after introduction into MEFs. *PUMA* shRNAs were capable of efficiently suppressing PUMA expression, even in *E1A/ras*-expressing cells that harbor stabilized p53 (Fig. 1*A*) (data not shown). Furthermore, like a p53 shRNA, *PUMA* shRNAs protected *E1A/ras*-expressing MEFs against apoptosis after serum depletion (Fig. 1*B*). Therefore, *PUMA* shRNAs are capable of modulating *PUMA* expression and activity.

PUMA Suppression Does Not Affect p53-Dependent Cell-Cycle Arrest. p53 is essential for cell-cycle arrest after DNA damage (24). To rule out the possibility that PUMA affects p53 arrest functions, we examined the ability of MEFs expressing PUMA or p53 shRNAs to undergo cell-cycle arrest after γ-irradiation. Both vector control and shPUMA-expressing MEFs retained an intact

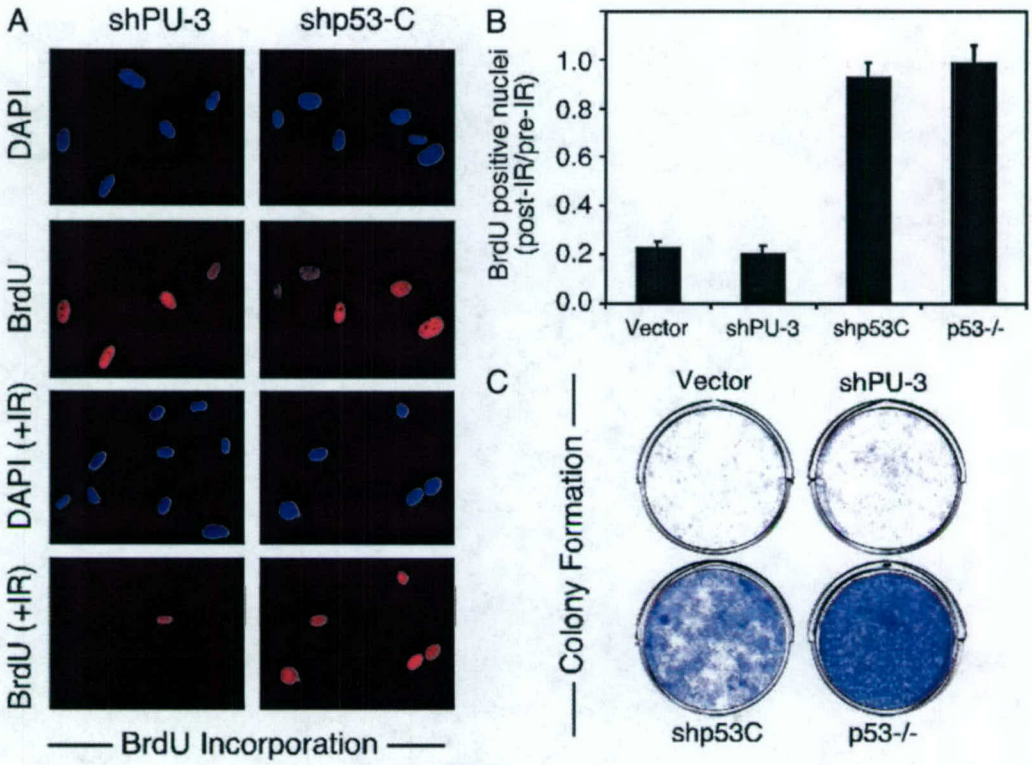


Fig. 2. *PUMA* suppression does not impair p53-mediated cell-cycle arrest. (*A*) Representative fields of a BrdUrd-incorporation assay showing the response of *p53* and *PUMA* shRNA-expressing MEFs to ionizing radiation. (*B*) Quantitation of the data in *A*, including MSCV control-infected and $p53^{-/-}$ MEFs. The relative ratios of BrdUrd incorporation for irradiated vs. nonirradiated cells are shown. (*C*) A colony-formation assay performed on $p53^{-/-}$ MEFs and WT MEFs infected with MSCV control (vector), shPU-3, and shp53. In each well, 2,500 cells were plated and cultured for 12 days before staining with crystal violet.

DNA damage checkpoint after irradiation, as indicated by an ≈5-fold decrease in BrdUrd relative to unirradiated controls (Fig. 2 *A* and *B*). In contrast, shp53-expressing MEFs and *p53*-null MEFs showed no significant reduction in BrdUrd incorporation after γ-irradiation (Fig. 2 *A* and *B*).

WT MEFs undergo senescence when plated at clonogenic density, whereas *p53*-deficient MEFs form colonies that are readily immortalized (19). To determine whether PUMA could influence the ability of p53 to promote senescence, MEFs expressing shRNAs targeting *PUMA* or *p53* were plated at low density and examined for colony formation after 2 weeks. In agreement with the acute arrest assays, *PUMA* shRNAs had no impact on colony formation, whereas p53 shRNAs were highly effective (Fig. 2*C*). In this assay, *p53*-deficiency and *p53* shRNAs resulted in the significant enhancement of the ability of untransformed cells to form colonies when plated at clonogenic density (Fig. 2*C*). Therefore, although suppression of *PUMA* can effectively disable p53 apoptotic functions, it has no impact on p53-mediated cell-cycle arrest.

PUMA Can Be a Potent Suppressor of Transformation. p53 tumor suppressor activity has been extensively studied in primary rodent cells, where *p53* mutations can cooperate with the combination of *E1A* and *ras* oncogenes, or oncogenic *ras* alone, in promoting oncogenic transformation (19, 25, 26). To determine the extent to which *PUMA* suppression could mimic *p53* mutations in these assays, we introduced shRNAs targeting *PUMA* or *p53* into WT or $p53^{-/-}$ MEFs, along with retroviruses that coexpressed *E1A/ras* or *ras* alone. After transduction, the infected cell populations were injected s.c. into immunocompromised mice, which were monitored for tumor formation at the sites of injection. Because the *PUMA* and *p53* shRNAs also coexpressed a GFP reporter, tumor formation also could be visualized by whole-body fluorescence imaging.

PUMA shRNAs acted as potent inducers of transformation in cells coexpressing *E1A/ras*, because the transduced cell populations appeared morphologically transformed (data not shown) and formed rapidly progressing tumors at the majority of injected sites (Fig. 3). Indeed, the ability of PUMA shRNAs to enhance the tumorigenicity of *E1A/ras* MEFs was similar to that produced by a *p53* shRNA (Fig. 3*A*), although the tumors progressed at a somewhat slower rate (Fig. 3*B*). Importantly, no tumors occurred in *E1A/ras* MEFs infected with the control vector. Furthermore, both *PUMA* shRNAs tested were effective in this assay, indicating that their oncogenic effects were unlikely to result from off-target oncogenic activities. Therefore, PUMA can approximate p53 action in suppressing transformation by *E1A* and *ras*.

The same transforming effects of *PUMA* shRNAs were not observed in the presence of oncogenic *ras* alone. Hence, cell populations expressing *PUMA* shRNAs and *ras* appeared morphologically senescent (data not shown) and, like cells coexpressing *ras* and the control vector, did not form tumors at any of the injected sites (Fig. 3 *A* and *C*). These observations are in stark contrast to WT MEFs expressing a *p53* shRNA or p53-deficient MEFs, where oncogenic *ras* was highly tumorigenic. Therefore, PUMA suppression can approximate p53 loss in promoting transformation by some oncogene combinations, but not others. Because both the *E1A/ras* and *ras* transformation assays were performed in the same MEF populations, these differences must reflect a distinct requirement for p53 effectors in different signaling environments.

***PUMA* Loss Accelerates *Eμ-myc* Lymphomagenesis.** *Eμ-myc* transgenic mice express the c-*myc* oncogene from an Ig heavy chain

GENETICS

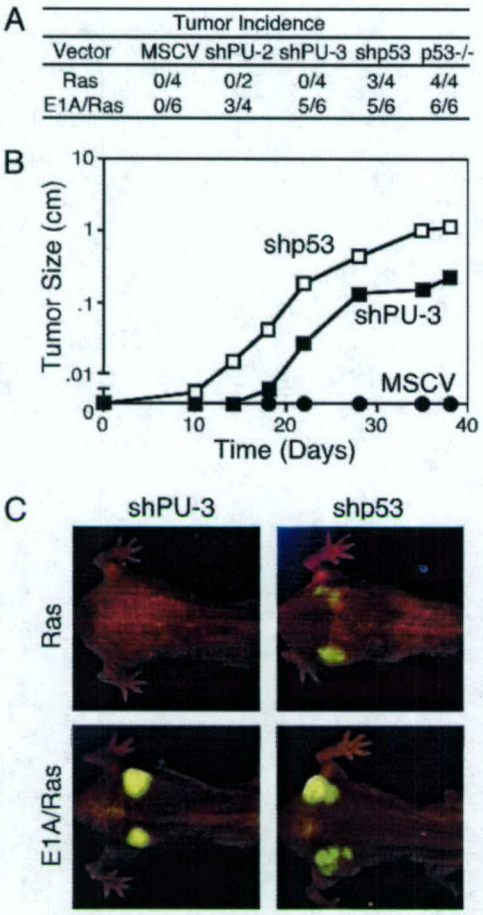


Fig. 3. *PUMA* loss transforms *E1A/ras* MEFs but not *ras*-transduced MEFs. (*A*) *E1A/ras* or *ras* MEFs coexpressing a control vector, PUMA shRNAs, or a p53 shRNA were injected s.c. into athymic nude mice and monitored for tumor formation. Shown is the number of tumors per injected site. The tumor incidence from $p53^{-/-}$ MEFs expressing *E1A/ras* or *ras* is also displayed. (*B*) Relative growth rate of s.c. tumors induced by MSCV-, shPU-3-, and shp53-infected *E1A/ras* MEFs. (*C*) GFP imaging of tumors arising from *ras* and *E1A/ras* MEFs transduced with *PUMA* and *p53* shRNAs.

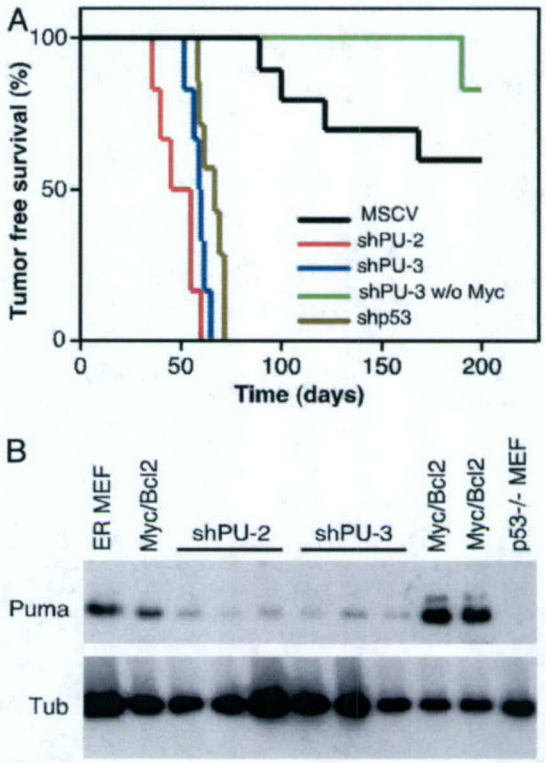


Fig. 4. Acceleration of *myc*-induced lymphomagenesis by *PUMA* shRNAs. (*A*) Mice reconstituted with stem cells infected with the indicated *PUMA* shRNAs were monitored for tumor onset and illness until they reached a terminal stage and were killed. The data are presented in a Kaplan–Meier format showing the percentage of mouse survival at various times postreconstitution. The shp53 survival data represents an updated cohort of mice, including previously published shp53-recipient mice (16). (*B*) Western blot of shPUMA-induced tumors showing decreased levels of PUMA relative to control *Eμ-myc bcl2* lymphomas and MEFs infected with *E1A/ras*. Tubulin is shown as a loading control.

enhancer and develop B cell lymphomas between 3 and 6 months of age (27). However, *Eμ-myc* lymphomas harboring *p53* deletions arise much more rapidly and typically display a more aggressive and disseminated pathology (28). Hematopoietic stem cells (HSCs) from *Eμ-myc* transgenic mice also give rise to lymphomas upon adoptive transfer into normal recipients, and these lymphomas can be greatly accelerated by *p53* suppression by RNAi (16). Given the ability of PUMA to mediate p53-dependent apoptotic activity in MEFs, we investigated whether stable suppression of PUMA in *Eμ-myc* hematopoietic stem cells by RNAi could recapitulate the effects of *p53* deletions and accelerate lymphomagenesis in recipient animals.

HSCs derived from *Eμ-myc* fetal livers were infected with retroviruses encoding a control vector, shPU-2, shPU-3, or a *p53* shRNA, and the resulting populations were used to reconstitute the hematopoietic system of lethally irradiated mice. Whereas only 40% of mice reconstituted with control HSCs developed lymphomas (four of 10, average survival of 120 ± 29 days), 100% of the mice receiving shPUMA-expressing HSCs developed lymphomas with dramatically reduced latency ($n = 12$, average survival 49 ± 7 and 59 ± 4 for shPU-2 and shPU-3, respectively; $P < 0.001$ relative to controls). Importantly, lymphomas arising from shPUMA-transduced shRNAs contained the shPUMA retrovirus and suppressed PUMA protein because all 12 of these lymphomas were GFP-positive (compared to only 1 of 4 controls) and showed substantially reduced PUMA expression relative to *E1A/ras* MEFs and *bcl2*-expressing lymphomas (Figs. 4*B* and 5*A*). Importantly, *PUMA* shRNAs were as potent as a *p53* shRNA in promoting lymphomagenesis (Fig. 4*A*) (16). Moreover, *myc* was required for these effects because only one-sixth of mice reconstituted with WT HSCs expressing *PUMA* shRNAs formed a lymphoma of T cell origin. Therefore, like *p53* loss, PUMA suppression can potently cooperate with *myc* during lymphomagenesis.

shPUMA Lymphomas Are Aggressive Pre-B Cell Lymphomas That Retain p53-Dependent Cell-Cycle Checkpoints. To further characterize tumorigenesis produced by *PUMA* suppression, we monitored lymphoma manifestation in recipient animals and conducted a variety of pathological analyses. In agreement with the survival data, whole-body fluorescence imaging of GFP expression in developing lymphomas revealed that the onset and overall distribution of shPUMA lymphomas closely resembled that occurring in shp53 lymphomas. Specifically, as was shown for *Eμ-myc* shp53 tumors in ref. 16, *Eμ-myc* shPUMA lymphomas typically involved many of the peripheral lymph nodes, including cervical, inguinal, brachial, and mesenteric lymphomas (Fig. 5*A*) (data not shown), with significant dissemination into the liver and spleen (Fig. 5*B*) (data not shown). Consistent with previous studies on *p53*-null lymphomas, both shPUMA and shp53 lymphomas rarely displayed the "starry sky" histology

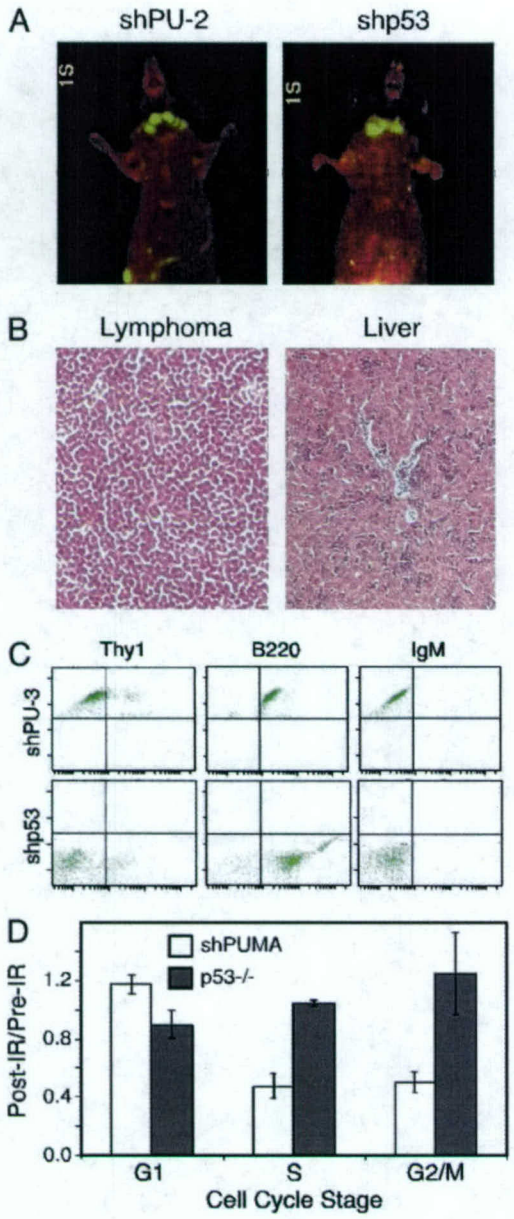


Fig. 5. shPUMA lymphomas phenocopy shp53 lymphomas. (*A*) *In vivo* GFP imaging showing cervical lymphomas in mice reconstituted with *Eμ-myc* HSCs transduced with PUMA and p53 shRNAs. (*B*) Hematoxylin/eosin staining of shPUMA lymphoma and liver sections, showing an absence of apoptotic cell clusters and perivascular and perenchymal infiltration of tumor cells, respectively. (*C*) Immunophenotyping of shPUMA and shp53 tumors by flow cytometry reveals that both tumors are pre-B cell lymphomas. Unlike the shPUMA vector, the shp53 vector lacks GFP expression. Thus, the resulting tumors are GFP-negative. (*D*) Mice harboring shPUMA and $p53^{-/-}$ lymphomas were left untreated or irradiated at 6 Gy, and DNA content analysis was performed on extracted lymphoma cells 36 h later. The relative ratios of the percent of irradiated vs. nonirradiated shPUMA and $p53^{-/-}$ lymphoma cells in each cell-cycle stage are shown.

indicative of the extensive apoptosis observed in *Eμ-myc* lymphomas (Fig. 5*B*) (6), and both were classified as pre-B cell lymphomas by immunophenotyping (Fig. 5*C*). Thus, the effects of p53 loss and *PUMA* suppression on *myc*-induced lymphomagenesis are strikingly similar.

The data described above suggest that disruption of apoptosis through *PUMA* suppression can mimic *p53* loss during *myc*-induced lymphomagenesis. To determine whether shPUMA lymphomas have acquired defects in p53-dependent arrest functions, we examined the integrity of the radiation-induced cell-cycle checkpoint in shPUMA lymphomas. Tumor-bearing mice were subjected to whole body γ-irradiation, and, 36 h later, the lymphomas were harvested and subjected to DNA content analysis by using flow cytometry. In contrast to $p53^{-/-}$ lymphomas but comparable with control and Bcl2-expressing lymphomas with intact p53 (8), shPUMA lymphomas accumulated in G_1 and showed a significant reduction in *S* phase after irradiation (Fig. 5*D*) (data not shown), implying that the effects of *PUMA* shRNAs on tumorigenesis do not require secondary mutations that compromise p53-dependent cell-cycle checkpoints. Interestingly, unirradiated p53-deficient lymphomas displayed a substantially increased S-phase population relative to shPUMA lymphomas at the time of isolation (data not shown). Therefore, although the increased proliferative rate associated with *p53* mutations may confer advantages over *PUMA* suppression at later stages of tumor evolution, our data indicate that loss of apoptosis through *PUMA* disruption phenocopies *p53* loss in promoting *Eμ-myc* lymphomagenesis.

Discussion

Our studies indicate that suppression of *PUMA* can approximate the effects of *p53* loss during *E1A/ras*-mediated transformation of primary MEFs and during *Myc*-induced lymphomagenesis. As such, they demonstrate that *PUMA* is an important component of the p53 tumor suppressor network and highlight the utility of stable RNAi technology to evaluate the activity of candidate tumor-suppressor genes.

These results demonstrate that *PUMA* can function as a *bona fide* tumor suppressor in mice. The effects of *PUMA* suppression were equivalent to *p53* loss in promoting lymphoma onset and were nearly as effective as *p53* loss during *E1A/ras*-induced transformation. Although mutations in BH3-only proteins have yet to be identified in human tumor specimens, our work predicts that *PUMA* suppression, through loss of p53 transactivating functions or direct mutations, may contribute to human cancer. Interestingly, *PUMA* maps to chromosome 19q13.3, which is altered in human gliomas, neuroblastomas, and B cell lymphomas (29–31). Notably, complete inactivation of *PUMA* may not be necessary to promote tumor phenotypes because, although *PUMA* expression was substantially reduced in shPUMA-expressing lymphomas, it was still detectable (Fig. 4*B*).

Interestingly, *PUMA*-null mice are not overtly tumor-prone (14, 15). Although, at first glance, these results are contradictory to our findings, they suggest that the ability of PUMA to mediate apoptosis and tumor suppression is context-dependent. In fact, our studies demonstrate that the ability of PUMA to act as a tumor suppressor can be dependent on other oncogenic events. Hence, although *PUMA* shRNAs cooperated effectively with *E1A/ras* and *myc* to promote tumorigenesis *in vivo*, they did not cooperate with *ras* alone. By analogy, *bcl2*, a PUMA antagonist, efficiently cooperates with *myc* during lymphomagenesis but is not a potent oncogene on its own.

Our study has important implications for understanding p53 action in tumor suppression. We previously showed that *bcl2* overexpression could mimic *p53* loss during *myc*-induced lymphomagenesis and suggested that disruption of apoptosis was sufficient to explain the tumorigenic advantage conferred by *p53* mutations in this context. Here, we show that PUMA, a p53 effector that promotes apoptosis but not cell-cycle arrest, can account for p53 action in at least some settings where p53 acts as a tumor suppressor. These data reinforce the notion that, despite the diversity of p53 activities, its action in tumor suppression can be mediated by only one effector function. Consequently, although *p53*-null lymphomas may have many defects, our data imply that some are byproducts of tumorigenesis and do not provide any immediate advantage to the developing tumor.

GENETICS

Importantly, PUMA does not approximate p53 action in modulating all tumor phenotypes. Whereas *p53* loss is an effective initiator of T cell lymphomagenesis (7), PUMA suppression is not, based on our results and the findings in refs. 14 and 15. Additionally, although *p53* and *PUMA* shRNAs transformed primary MEFs with *E1A/ras*, only *p53* shRNAs or *p53* deletions cooperated with *ras*-induced transformation. This context dependence can be understood in light of the underlying p53 biology. The *E1A* and *myc* oncogenes induce proliferation but also activate p53 to promote apoptosis (26, 32, 33). Hence, disruption of the p53-dependent apoptotic activity provides an immediate and potent advantage to the oncogene-expressing cells but provides little benefit to proliferation-restricted normal cells not subject to an apoptotic stimulus. In contrast, oncogenic *ras* promotes proliferation but also activates p53 to promote cellular senescence (20). Here, p53's arrest functions may be crucial for its tumor suppressor activity. Because PUMA does not mediate these functions, its inactivation does not mimic *p53* loss in this setting. The differential relevance of apoptosis in p53-mediated tumor suppression also may explain why certain *p53* mutants, defective in apoptosis but not cell-cycle arrest, are impaired in their ability to initiate T cell lymphomagenesis (10).

In summary, our results imply that the p53 functions underlying its tumor-suppressor activity are context-dependent and may be influenced by cell type, microenvironment, and oncogenic events acquired during the course of tumor evolution. In some settings, such as in *E1A/ras* transformation and *myc*-induced lymphomagenesis, disruption of apoptosis by *PUMA* loss or defects in other apoptotic regulators is sufficient to promote tumorigenesis, whereas defects in cell-cycle checkpoints and genomic instability are apparently byproducts of *p53* loss. However, these byproducts may provide new capabilities that become relevant later during tumor progression or cancer therapy, producing tumors that are more aggressive than those acquiring strictly antiapoptotic lesions (17). It is noteworthy that *p53*-deficient lymphomas, once established, progress to a lethal stage more rapidly than those expressing *PUMA* shRNAs, perhaps because of their higher proliferative capacity or increased genomic instability (data not shown). Undoubtedly, in other settings, defects in cell-cycle checkpoints provide the driving force for *p53* mutations, with apoptotic defects being byproducts of *p53* loss. Because essential p53 tumor-suppressor functions are context-dependent, effective strategies to treat *p53* mutant tumors also may depend on context. Understanding which p53 function(s) are key to the evolution of different tumor types may ultimately identify activities required for tumor maintenance and suggest targets for therapeutic intervention.

We thank L. Bianco for histology; R. Sachidanandam for bioinformatics support; C. Rosenthal, M.-M. Yang, and S. Ray for technical assistance; and R. Dickins and other members of the Lowe laboratory for advice and critical reading of the manuscript. This work was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation (to M.T.H.), a postdoctoral training grant (to J.T.Z.), National Cancer Institute Project Grants CA13106 (to G.J.H. and S.W.L.) and CA87497 (to S.W.L.), and a generous gift from the Laurie Strauss Leukemia Foundation. D.J.B. is the recipient of a Watson School of Biological Sciences Engelhorn Fellowship, and S.W.L. is an American Association for Cancer Research–National Foundation for Cancer Research Research Professor.

1. Vogelstein, B., Lane, D. & Levine, A. J. (2000) *Nature* **408,** 307–310.
2. Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. (1999) *Genes Dev.* **13,** 2658–2669.
3. Yin, C., Knudson, C. M., Korsmeyer, S. J. & Van Dyke, T. (1997) *Nature* **385,** 637–640.
4. Zong, W. X., Lindsten, T., Ross, A. J., MacGregor, G. R. & Thompson, C. B. (2001) *Genes Dev.* **15,** 1481–1486.
5. Fridman, J. S. & Lowe, S. W. (2003) *Oncogene* **22,** 9030–9040.
6. Schmitt, C. A., Fridman, J. S., Yang, M., Baranov, E., Hoffman, R. M. & Lowe, S. W. (2002) *Cancer Cell* **1,** 289–298.
7. Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Jr., Butel, J. S. & Bradley, A. (1992) *Nature* **356,** 215–221.
8. Strasser, A., Harris, A. W. & Cory, S. (1991) *Cell* **67,** 889–899.
9. Sentman, C. L., Shutter, J. R., Hockenbery, D., Kanagawa, O. & Korsmeyer, S. J. (1991) *Cell* **67,** 879–888.
10. Liu, G., Parant, J. M., Lang, G., Chau, P., Chavez-Reyes, A., El-Naggar, A. K., Multani, A., Chang, S. & Lozano, G. (2004) *Nat. Genet.* **36,** 63–68.
11. Yu, J., Zhang, L., Hwang, P. M., Kinzler, K. W. & Vogelstein, B. (2001) *Mol. Cell* **7,** 673–682.
12. Nakano, K. & Vousden, K. H. (2001) *Mol. Cell* **7,** 683–694.
13. Yu, J., Wang, Z., Kinzler, K. W., Vogelstein, B. & Zhang, L. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 1931–1936.
14. Villunger, A., Michalak, E. M., Coultas, L., Mullauer, F., Bock, G., Ausserlechner, M. J., Adams, J. M. & Strasser, A. (2003) *Science* **302,** 1036–1038.
15. Jeffers, J. R., Parganas, E., Lee, Y., Yang, C., Wang, J., Brennan, J., MacLean, K. H., Han, J., Chittenden, T., Ihle, J. N., *et al.* (2003) *Cancer Cell* **4,** 321–328.
16. Hemann, M. T., Fridman, J. S., Zilfou, J. T., Hernando, E., Paddison, P. J., Cordon-Cardo, C., Hannon, G. J. & Lowe, S. W. (2003) *Nat. Genet.* **33,** 396–400.
17. Schmitt, C. A., Fridman, J. S., Yang, M., Lee, S., Baranov, E., Hoffman, R. M. & Lowe, S. W. (2002) *Cell* **109,** 335–346.
18. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. (2002) *Genes Dev.* **16,** 948–958.
19. Lowe, S. W., Jacks, T., Housman, D. E. & Ruley, H. E. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2026–2030.
20. Serrano, M., In, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. (1997) *Cell* **88,** 593–602.
21. Zilfou, J. T., Hoffman, W. H., Sank, M., George, D. L. & Murphy, M. (2001) *Mol. Cell. Biol.* **21,** 3974–3985.
22. Lin, A. W., Barradas, M., Stone, J. C., van Aelst, L., Serrano, M. & Lowe, S. W. (1998) *Genes Dev.* **12,** 3008–3019.
23. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. (1993) *Cell* **74,** 957–967.
24. Kastan, M. B., Zhan, Q., el-Deiry, W. S., Carrier, F., Jacks, T., Walsh, W. V., Plunkett, B. S., Vogelstein, B. & Fornace, A. J., Jr. (1992) *Cell* **71,** 587–597.
25. Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. & Rotter, V. (1984) *Nature* **312,** 649–651.
26. Debbas, M. & White, E. (1993) *Genes Dev.* **7,** 546–554.
27. Adams, J. M., Harris, A. W., Pinkert, C. A., Corcoran, L. M., Alexander, W. S., Cory, S., Palmiter, R. D. & Brinster, R. L. (1985) *Nature* **318,** 533–538.
28. Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. (1999) *Genes Dev.* **13,** 2670–2677.
29. Shimazaki, C., Inaba, T. & Nakagawa, M. (2000) *Leuk. Lymphoma* **38,** 121–130.
30. Yong, W. H., Chou, D., Ueki, K., Harsh, G. R., IV, von Deimling, A., Gusella, J. F., Mohrenweiser, H. W. & Louis, D. N. (1995) *J. Neuropathol. Exp. Neurol.* **54,** 622–626.
31. Mora, J., Cheung, N. K., Chen, L., Qin, J. & Gerald, W. (2001) *Clin. Cancer Res.* **7,** 1358–1361.
32. Lowe, S. W. & Ruley, H. E. (1993) *Genes Dev.* **7,** 535–545.
33. Rao, L., Debbas, M., Sabbatini, P., Hockenbery, D., Korsmeyer, S. & White, E. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7742–7746.

Unlocking the potential of the human genome with RNA interference

Gregory J. Hannon[1] & John J. Rossi[2]

[1]Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA (email: hannon@cshl.edu) [2]Division of Molecular Biology, Beckman Research Institute of the City of Hope, Graduate School of Biological Sciences, Duarte, California 91010, USA (email: JRossi@coh.org)

The discovery of RNA interference (RNAi) may well be one of the transforming events in biology in the past decade. RNAi can result in gene silencing or even in the expulsion of sequences from the genome. Harnessed as an experimental tool, RNAi has revolutionized approaches to decoding gene function. It also has the potential to be exploited therapeutically, and clinical trials to test this possibility are already being planned.

The formal description of RNAi as a biological response to double-stranded RNA (dsRNA) came about following experiments with dsRNA in the nematode *Caenorhabditis elegans*[1,2]. Injecting dsRNAs into the worm was found to silence genes whose sequences were complementary to those of the introduced dsRNAs[3]. It is now clear that an RNAi pathway is present in many, if not most, eukaryotes[4]. dsRNAs are processed into short interfering RNAs (siRNAs), about 22 nucleotides in length, by the RNase enzyme Dicer. These siRNAs are then incorporated into a silencing complex called RISC (RNA-induced silencing complex), which identifies and silences complementary messenger RNAs.

Before RNAi could be harnessed as an experimental tool for silencing specific genes in mammalian systems, a considerable hurdle had to be overcome. The problem lay in making exogenous dsRNA trigger silencing in a gene-specific manner without invoking nonspecific responses to foreign dsRNAs that are part of the cell's antiviral mechanism[5]. Work over the past two years has allowed investigators to meet this challenge, and RNAi has now been adopted as a standard methodology for silencing the expression of specific genes in mammalian cells. Here, we chronicle the development of RNAi as a genetic tool in mammals, focusing on recent advances and providing practical advice for its experimental application. We also make predictions about the potential future of RNAi as a potent and specific therapeutic tool that may escape some of the limitations of conventional medicinal chemistry.

Breaking the barrier to RNAi in mammals

For more than 30 years, it has been known that exposure of mammalian cells to long dsRNAs induces innate immune pathways, including interferon-regulated responses that serve as antiviral mechanisms. The enzyme dsRNA-dependent protein kinase (PKR) is activated on binding to dsRNA and localized, but sequence-independent destruction of RNAs and a generalized repression of protein synthesis results[5]. The existence of these innate immune pathways seemed incompatible with the use of dsRNA for silencing a particular target gene. However, evidence of an RNAi pathway in mammals came from the observation that key biochemical components of the RNAi pathway are conserved[6,7]. It was also shown that long dsRNAs can trigger gene-specific responses when they are introduced into mammalian embryos and embryonic cell lines in which nonspecific antiviral responses to dsRNAs are not prevalent[4]. This raised the problem of how to shift the response of a mammalian cell to foreign dsRNA from the non-specific sequence-independent defence pathways to the sequence-specific RNAi pathway. Attempts to meet this challenge have resulted in RNAi being established as a genetic tool in mammalian cells and animals.

Using siRNAs for RNAi

A biochemical understanding of the RNAi pathway (Fig. 1; see review in this issue by Meister and Tuschl, page 343) was crucial to realizing that dsRNAs shorter than 30 base pairs (bp) could be used to trigger an RNAi response in mammals. Tuschl and colleagues showed that transfection of mammalian cells with short RNAs could induce the sequence-specific RNAi pathway, and so overcame the barrier to the use of RNAi as a genetic tool in mammals[8]. The impetus to use siRNAs and other small RNAs in mammalian cells also came from the long-standing view that PKR activation and similar responses were not effectively triggered by short dsRNAs. Following the initial reports, it took a remarkably short period of time for siRNAs triggers to be adopted as a standard component of the molecular biology toolkit.

siRNAs can be introduced into mammalian cells using a variety of standard transfection methods. The strength and duration of the silencing response is determined by several factors: on a population basis, the silencing response is affected mainly by the overall efficiency of transfection, which can be addressed by optimizing conditions. In each cell, silencing depends on the amount of siRNA that is delivered and on the potential of each siRNA to suppress its target, or its potency. Even a relatively impotent siRNA can silence its target provided that sufficient quantities of the siRNA are delivered. However, essentially 'forcing' the system by delivering large amounts of reagent is likely to lead to numerous undesired effects (see section 'Intrinsic limits on the specificity of RNAi').

Using shRNAs for RNAi

The discovery of the endogenous triggers of the RNAi pathway in the form of small temporal RNAs — now termed microRNAs (miRNAs)[9–11] — suggested that RNAi might be triggered in mammalian cells by synthetic genes that express mimics of endogenous triggers. Several laboratories simultaneously used related approaches to test this idea. These involved expressing mimics of miRNAs in the form of short hairpin RNAs (shRNAs) from RNA polymerase II or III promoters[12]. The shRNAs themselves varied in size and design, with stems ranging from 19 to 29 nucleotides in length, and with various degrees of structural similarity to natural miRNAs. All these approaches were

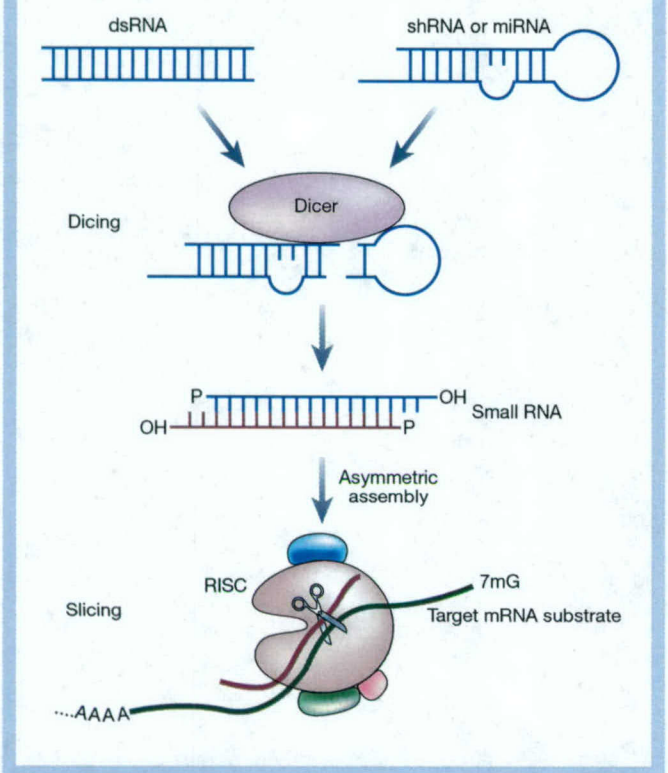


Figure 1 RNA silencing pathways in different organisms. Long dsRNA and miRNA precursors are processed to siRNA/miRNA duplexes by the RNase-III-like enzyme Dicer. These short dsRNAs are subsequently unwound and assembled into effector complexes, RISCs, which can direct RNA cleavage, mediate translational repression or induce chromatin modification. *S. pombe*, *C. elegans* and mammals carry only one Dicer gene. In *D. melanogaster* and *A. thaliana*, specialized Dicer or DLC proteins preferentially process long dsRNA or miRNA precursors. 7mG, 7-methyl guanine; AAAA, poly-adenosine tail; Me, methyl group; P, 5′ phosphate.

effective to varying degrees, and at present, no real consensus has developed on the most effective way to present synthetic miRNAs to the RNAi pathway.

Because these triggers are encoded by DNA vectors, they can be delivered to cells in any of the innumerable ways that have been devised for delivery of DNA constructs that allow ectopic mRNA expression. These include standard transient transfection, stable transfection and delivery using viruses ranging from retroviruses to adenoviruses. Expression can also be driven by either constitutive or inducible promoter systems[12].

Recent studies strongly indicate that each shRNA expression construct gives rise to a single siRNA (D. Siolas, G.J.H. and M. Cleary, unpublished work). Knowing precisely how this processing occurs for each shRNA cassette design has permitted the application of siRNA design criteria (see 'Features of effective siRNAs and shRNAs' below) to the design of effective shRNAs. The use of siRNAs and shRNAs are complementary approaches in the application of RNAi as a genetic tool in mammals, and the best approach depends on the type of study being performed.

Features of effective siRNAs and shRNAs

Observations of widely varying efficacy of individual siRNAs motivated a search for rules that might specify more effective siRNAs. Several groups took a 'black box' approach, which involved assaying large numbers of siRNAs, sorting them into classes depending on their potency, and then looking for characteristics that distinguished effective siRNAs from ineffective ones[13–15]. Some common rules have begun to emerge from these studies. siRNAs in which the helix at the 5′ end of the antisense strand has a lower stability than the 3′ end of the siRNA are generally more effective silencers than those with the opposite arrangement. A biochemical basis for the thermodynamic arrangement of effective siRNAs was provided by biochemical studies of the mRNA-cleavage complex, RISC (Fig. 1), in *Drosophila* embryo extracts. These studies showed unequal incorporation of the two strands of the siRNA into RISC[14]. Moreover, strand biases could be manipulated by altering the thermodynamic stability of the terminal nucleotides in a way that precisely matched the rules derived from empirical studies. Finally, an examination of miRNAs, most of which produce RISC-like complexes containing only one strand of the precursor (see review in this issue by Meister and Tuschl, page 343), showed the same pattern of thermodynamic asymmetry as that shown by effective siRNAs[13,14,16,17]. The rules for specifying effective siRNAs uncovered by these studies imply that the effectiveness of an RNAi response triggered by an siRNA is strongly dependent on siRNA structure and determined at the step of RISC assembly, during which the asymmetry in the dsRNA must be sensed and a single strand chosen for productive incorporation into the enzyme. Once the active RISC is formed, it is relatively insensitive to the placement or structure of the target site within the mRNA.

Intrinsic limits on the specificity of RNAi

Although RNAi silences gene expression in a sequence-specific manner, several recent studies have suggested that the specificity of silencing is not absolute. Off-target effects in mammals can come from several different sources. As discussed previously, transfection of cells with dsRNAs can activate innate immune pathways. PKR activation was thought to depend on the length of the dsRNA, with a minimal cut-off for PKR activation being roughly 30 bp of duplex. However, recent reports have suggested that both siRNAs and shRNAs can — under some circumstances and in certain cell types — activate a PKR response[18–20]. Furthermore, siRNAs transcribed *in vitro* using bacteriophage polymerases can be potent activators of an interferon response if the initiating triphosphate is not completely removed from the transcripts[21]. Further studies are required to investigate the frequency with which RNAi triggers provoke these antiviral response pathways, and the sequence or structural characteristics that might lead an siRNA or an shRNA to trigger such a response.

miRNAs recognize and regulate their targets despite a lack of perfect complementarity. This raises the possibility that siRNAs might also not require contiguous base pairing to suppress their targets effectively; several microarray studies suggested that siRNAs can provoke sequence-dependent, off-target effects and that these can be elicited by 14 base pairings, or possibly even fewer, between the siRNA and its target[22]. Notably, analysis of such interactions suggested that base pairing at the 5′ end of the siRNA contributed disproportionately to targeting, a conclusion also reached by analysis of interactions between miRNAs and their validated targets[23–26]. Although such information can aid the design of more specific siRNAs, we do not have sufficient understanding of target recognition by RISC to say with certainty that we can eliminate off-target effects. In fact, the intrinsic specificity of the RNAi pathway may be sufficiently low to prevent the design of a completely specific siRNA in mammals. Fewer studies have been carried out with shRNA expression cassettes, but similar caveats undoubtedly apply.

Off-target effects can also occur at the level of protein synthesis. miRNAs in animals often regulate protein expression without having corresponding effects on mRNA levels. Several studies have indicated that siRNAs also do this, provided that mRNA cleavage is blocked by altering the geometry of the target–substrate interaction[24–26]. However, suppression of a reporter gene in this manner was only effective when several siRNA binding sites were present. Although these findings might provide some degree of comfort to those using siRNAs experimentally, a recent study suggested that changes in the expression levels of a large number of proteins occurred in cells treated

with siRNAs. However, it should be noted that the siRNAs in question were relatively impotent[27]. Therefore, we must view unwanted changes in protein expression levels as the 'monster under the bed' for RNAi-based studies of gene function. Ultimately, caveats in the specificity of the RNAi response make it essential to follow relatively simple guidelines for good experimental practice. These are outlined in Box 1.

RNAi as a solution for mammalian genetics

One of the first choices to make in any RNAi-based genetic experiment is whether to trigger suppression through the use of siRNAs or shRNAs. The advantages of using siRNAs are relative ease of availability and high efficiency of delivery. In addition, pre-validated siRNAs are becoming increasingly available as commercial suppliers and the scientific community acquire more experience. Overall, siRNA delivery is likely to result in the highest intracellular concentration of the gene silencer. But limitations to the use of siRNAs are that their effects are transient and restricted by the rate of cell division: mammalian cells do not have mechanisms to amplify and propagate RNAi (unlike *C. elegans* and plants). In addition, some cell types are notoriously difficult to transfect, and the procedure of transfection itself can alter the physiology of the cell. However, despite these drawbacks, transfection of siRNAs is probably the fastest and easiest method currently available for producing a knockdown of gene expression in cell culture by means of RNAi.

With shRNAs, the up-front investment is greater. First, DNA oligonucleotides must be cloned and sequenced so that a construct can be produced. Second, the shRNA must be designed effectively, and consensus on the most effective design, with respect to either the structure of the shRNA itself or the structure of the expression vector, is only just beginning to emerge. However, shRNAs are capable of producing sustained repression, and allow for delivery by conventional transfection or by several advanced viral vectors that also permit stable integration into the genome. In addition, shRNA expression vectors can be propagated indefinitely. As with siRNAs, design algorithms can be applied to shRNAs to maximize the probability of success in a suppression experiment. However, the application of such algorithms requires a detailed understanding of the vector system being used.

Both siRNAs and shRNAs have been used for studies of gene function *in vivo*, primarily in mice. Both types of trigger can be

Box 1

Rules of the road for effective RNAi experiments

Given the significant concerns about the specificity of RNAi-mediated repression, how can investigators maximize confidence in the results of studies that use these tools? It is important to note that no approach used to inactivate gene function is free from potential problems. Even conventional gene knockouts are known to be subject to compensation during development. Thus, the enthusiasm for the use of RNAi as a genetic tool should be tempered by a recognition of the potential problems and good practices should be applied to avoid misinterpreting results. Four guidelines for good practice in RNAi experiments in mammals are presented below.

1. Get the right strand into RISC by using good design

RNAi-based experiments will be more informative and go more smoothly if effective and highly specific RNA triggers are used. Many algorithms now exist for choosing effective sequences. In addition, homology to other sequences in the genome should be minimized, with particular attention to the 5′ end of the antisense strand. Use of design algorithms based on thermodynamic criteria can aid biased incorporation of the antisense strand of the siRNA into RISC. Several public websites provide support for such designs (see for example http://web.mit.edu/mmcmanus/www/home1.2files/siRNAs.htm; http://hydra1.wistar.upenn.edu/Projects/siRNA/siRNAindex.htm; http://www.cshl.edu/public/SCIENCE/hannon.html).

2. Several alleles are better than one

Several siRNAs or shRNAs should give the same phenotypic outcome, as it is extremely unlikely that different triggers will have the same off-target effects[22]. It is critical to correlate this phenotypic outcome with the effectiveness of suppression. Only effective siRNAs against a given target, but not inefffective siRNAs, should yield similar phenotypes. Importantly, discrimination between effective and ineffective siRNAs can only be accomplished by examining target protein levels. There are numerous anecdotal reports of siRNAs effectively suppressing protein production without changing mRNA levels. In addition, siRNAs or shRNAs that do not affect the target protein should be used as negative controls. Arguably, one could use a 'scrambled' siRNA or shRNA for this purpose. However, such scrambled siRNAs may not have any biological activity, and it is undoubtedly better to use an RNA that is known to enter the RNAi pathway effectively. For example, an RNA targeting luciferase, green fluorescent protein or another marker gene (that is validated against its target) would be expected to enter RISC but would not be expected to affect the expression of proteins in a mammalian cell. Other possible controls include an RNA with flipped asymmetry. This could be achieved by creating an siRNA with a more stable helix at the 5′ end of the antisense strand.

3. Work at the lowest possible concentrations

RISC is a conventional enzyme, and working at high enzyme to substrate ratios is likely to affect its specificity. Therefore, it is important to identify RNAi triggers that work at very low effective concentrations. With siRNAs, this can be achieved by titrating siRNA concentrations and by correlating their effects on phenotypic outcome with both the concentration of the siRNA used and with the degree of suppression obtained. For example, if the siRNA shows maximal suppression at 5 nM but the phenotype is not observed until the concentration reaches 100 nM, off-target effects must be suspected. In fact, some siRNAs in HeLa cells have shown IC_{50} values (the amount of siRNA required to suppress the target to 50% of its original level) of as little as 500 pM. Similarly, titration of shRNA-expression vectors should also be performed.

4. Rescue to the rescue

Ultimately, the best experiments demonstrate that expression of a version of the targeted gene that cannot be recognized by the siRNA reverts the phenotype. This can be achieved in several ways. First, mutations can be introduced into a cDNA encoding the targeted gene that destroy complementarity with the siRNA or shRNA while maintaining the wild-type protein sequence. Alternatively, the phenotype can be validated by using siRNAs or shRNAs that target untranslated regions, and then by rescuing the phenotype with an expression construct containing only the coding sequence. Although rescue experiments provide the ultimate test of the specificity of a given effect, these can be problematic. For example, it may be difficult to achieve appropriate expression levels of a particular protein. Overexpression could cause artefactual effects (for example, a pathway could be rescued by bypassing its requirements, rather than truly reverting a specific effect).

Ultimately, as our understanding of the RNAi pathway deepens, we will be able to predict with good accuracy all the on- and off-target effects of siRNAs. This will allow not only the generation of RNAi triggers with maximal specificity, but also the design of triggers that are directed against the most likely off-target genes for each siRNA or shRNA.

delivered transiently. The first demonstration of RNAi-mediated repression in an adult animal showed effective repression of a luciferase reporter gene following hydrodynamic transfection of siRNAs or shRNA expression plasmids into mouse liver[28]. Subsequent studies have delivered siRNAs or shRNAs by various methods, including lipid-based delivery and naked RNA or DNA injection[29–32].

Long-term gene silencing has been demonstrated *in vivo* using both genetic mosaics and germline modification. For example, the growth of a tumour cell line in a xenograft model can be attenuated by engineering that cell line with an shRNA cassette that targets the activated *ras* oncogene before the tumour cells are subcutaneously injected into the host animal[33]. Genetically mosaic animals have been created by engineering stem cells with shRNA vectors and then by using those stem cells to repopulate an organ system[34,35]. Strains of mice have been engineered to heritably suppress a targeted gene based on inheritance of a dominantly acting shRNA expression cassette[36–39]. Several approaches have been used to create such strains, including standard nuclear injection, creation of chimaeras with engineered embryonic stem cells, and transgenesis mediated by subzonal injection of fertilized eggs with recombinant lentiviruses. Ultimately these developments will rapidly lead to the creation of animals with inducible, tissue-specific silencing of almost any gene. RNAi is therefore likely to complement existing large-scale efforts to functionally map the mouse genome by chemical or insertional mutagenesis. RNAi is certainly complementary to such approaches, because each approach can generate different types of allele. However, unlike mutational approaches RNAi has the potential to be extended beyond mice into animals where recombinant organisms cannot be generated using embryonic stem cells.

RNAi as a tool for genome-wide studies

The success in using RNAi for analysing single genes has led inevitably to efforts to apply this approach on a large scale for forward genetics (whereby mutant genes are isolated from organisms showing abnormal physical and behavioural characteristics). Indeed, given the recent completion of the human, mouse and rat genomes, RNAi provides a ready mechanism by which this enormous wealth of sequence information can be translated into functional definitions for every gene.

Genome-wide libraries of siRNAs can be constructed in fundamentally different ways, including chemical synthesis or enzymatic digestion of long dsRNAs. An example of progress towards this goal can be found in a small-scale effort[40] to target genes in the phosphatidylinositol 3-OH kinase (PI(3)K) pathway in which a mini-library (148 siRNAs) was searched for genes that affected phosphorylation of Akt, a downstream substrate for PI(3)K.

Alternatively, libraries can be produced by constructing shRNA expression vectors that target each gene. As with siRNAs, proof of principle came initially from small-scale efforts. Using a library directed against the family of de-ubiquinating enzymes, the tumour suppressor CYLD (encoded by the familial cylindromatosis susceptibility gene) was identified as a suppressor of NF-κB activity[41]. This result led to proposals for treating cylindromatosis with existing drugs and provided powerful confirmation that unbiased, genetic approaches can lead not only to new insights in biology but also to practical advances in the treatment of disease. Two groups have recently reported the production of arrayed libraries from chemically synthesized oligonucleotides that cover about 10,000 different human genes each[42,43]. Another group has generated a library of polymerase chain reaction (PCR) products that encode shRNAs[44], and several groups have reported methods for constructing random shRNA libraries based on manipulation of complementary DNA or genomic DNA[45–47]. Each approach has specific advantages. Random libraries are relatively inexpensive to produce and can cover an individual gene with many different shRNAs. However, they suffer from a lack of normal representation of all genes. Libraries produced from chemically synthesized oligonucleotides are expensive. However, they permit the use of powerful informatic tools to aid shRNA design and allow flexibility in optimizing the structure of the shRNA for entry into the RNAi pathway. In addition, synthetic libraries can be used either as mixtures or as individual arrays, in a similar way to siRNA libraries.

Large-scale screening using siRNA libraries must be carried out by individual transfection and phenotypic characterization of target cells (Fig. 2). As such, siRNA libraries can be applied to the wide range of screening methods that are being developed by the pharmaceutical industry in the form of cell-based assays for drug development. These include fluorescent reporter screens, assays for various activities in cell lysates and screening by means of automated microscopy. Alternatively, RNAi triggers can be printed on microarrays and tested for their effects following transfection *in situ*[48,49]. Arrayed libraries of shRNAs can be used in a similar fashion: this was demonstrated by applying an arrayed library to a search for genes that affect proteasome function[43]. shRNA libraries can also be assayed, following their integration into the genomes of target cells, in pools using protocols that filter populations based on phenotypic criteria, such as a growth selection (Fig. 2). Such a test of one shRNA library yielded new links between several genes and the p53 tumour-suppressor pathway[42]. A conceptually more complicated application of pooled screens involves using molecular 'barcodes' to track how individual shRNAs behave as members of complex populations (Fig. 2).

Clearly, large-scale library efforts will evolve with our advancing understanding of the RNAi pathway. As the quality of resources improves, there will be opportunities to progress from relatively straightforward screening protocols in cultured cells to more complex genetics in whole animals.

RNAi in drug discovery and disease therapy

RNAi has begun to produce a paradigm shift in the process of drug discovery. With the large-scale screening approaches described above, RNAi can winnow lists of potential drug targets so that efforts can be focused on the most promising candidates. Moreover, since the first description of RNAi in mammalian cells, there have been numerous studies aimed towards using RNAi to treat disease. The strong appeal of RNAi in therapeutics is the potency and specificity with which gene expression can be inhibited. The possible targets for various diseases range from oncogenes to growth factors and single nucleotide polymorphisms (SNP). There is also potential for using RNAi for the treatment of viral diseases such as those caused by the hepatitis C virus (HCV) and the human immunodeficiency virus (HIV). Despite the excitement and some early proofs of principle in the literature, there are important issues and concerns about the therapeutic application of this technology, including difficulties with delivery and uncertainty about potential toxicity. However, proposals for clinical trials using either synthetic siRNAs or viral-vector-delivered shRNAs have been put forward — although none has yet been approved.

RNAi as a treatment for HIV

The development and use of double and triple drug combinations for the treatment of HIV infection has led to dramatic improvements in the lives of HIV-infected individuals. But despite the apparent successes of the new anti-retroviral drugs there are the emerging problems of drug-resistant viral variants and toxicities of the combination drugs now in use. Therefore, there is still great interest in exploring new antiviral therapeutic approaches. HIV was the first infectious agent targeted by RNAi, perhaps because the lifecycle and pattern of gene expression of HIV is well understood. Synthetic siRNAs and expressed shRNAs have been used to target several early and late HIV-encoded RNAs in cell lines and in primary haematopoietic cells including the TAR element[50], tat[51–53], rev[51,52], gag[54,55], env[55], vif[50], nef[50], and reverse transcriptase[53].

Despite the success of RNAi-mediated inhibition of HIV-encoded RNAs in cell culture, targeting the virus directly represents a substantial challenge for clinical applications because the high viral mutation rate

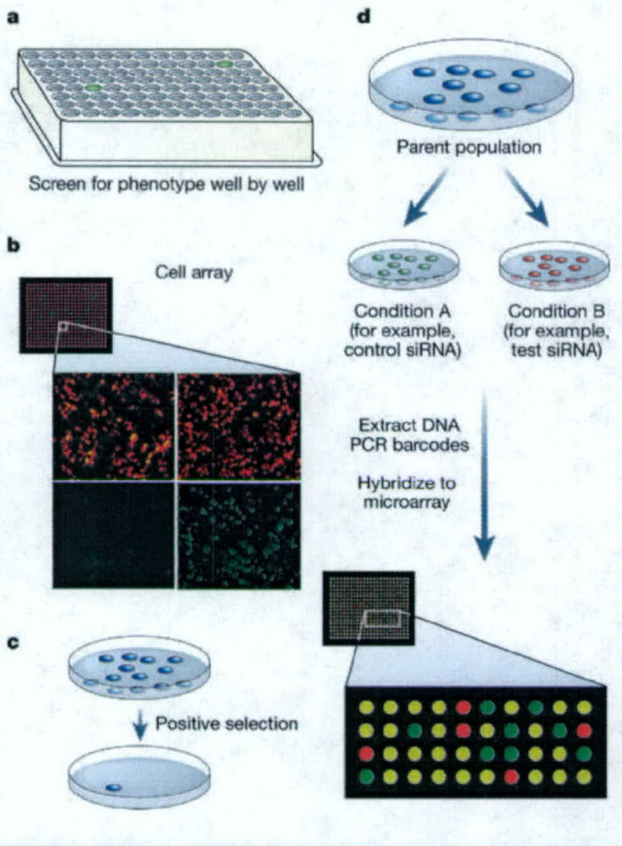


Figure 2 Genome-wide screens using RNAi. **a**, Standard methodologies can be used to screen siRNAs or shRNAs individually in 96-well plates using morphological readouts, reporters or biochemical assays. **b**, Similar approaches can be taken using siRNAs or shRNAs that have been printed on microarrays for reverse transfection[49]. Reverse transfection involves the deposition of lipid–nucleic-acid complexes on a solid surface, often a glass microarray slide. Cells plated on top of the slide take up the encapsulated DNA or RNA, and this can direct mRNA expression or gene silencing. The enlarged image shows cell populations, which would be observed within individual spots of the array, expressing fluorescent proteins, red and green. In the left panels, expression of GFP has been ablated by co-deposition of a GFP siRNA. **c**, Complex (mixed) populations of shRNA-expression vectors can be filtered through positive selections in cultured cells. Selection can be for drug resistance, cytokines, genetic alterations, or — in the case of fluorescence activated cell sorter (FACS)-based selection — for cells that activate a particular marker. **d**, Complex populations can be monitored using molecular barcodes that track individual shRNA-expressing cells and their responses to various stimuli. The use of molecular barcodes combines the advantages of well-by-well screens with the advantages of carrying out pooled selections, thereby allowing the identification of phenotypes in complex populations, which do not necessarily confer a growth advantage (for example, a synthetic lethal phenotype). Each vector is tagged with a unique sequence, which can comprise the shRNA itself or a separate random or selected barcode. The frequency (representation) of each vector in a mixed population can be measured by hybridizing barcodes to an oligonucleotide microarray. If the population is subjected to selective pressure, the representation of individual shRNA constructs is expected to change as a result. This change can be detected by comparing hybridization signals for the starting population with those of the population exposed to selection. The relative signal of shRNAs that increase resistance to the selection will increase, whereas the relative signal of those that sensitize to the selection will decrease.

will lead to mutants that can escape being targeted[56]. Therefore RNAi-mediated downregulation of the cellular cofactors required for HIV infection is an attractive alternative or complementary approach. Cellular cofactors such as NF-κB[53], the HIV receptor CD4[54], and the co-receptors CXCR4 and CCR5[57] have been successfully downregulated by RNAi, resulting in the inhibition of HIV replication in numerous human cell lines and in primary cells including T lymphocytes and haematopoietic-stem-cell-derived macrophages[50–52,54,57–60]. Although targeting NF-κB is not appropriate as a therapy owing to the important role NF-κB has in the cell (for example, it mediates interferon-induced gene expression) the macrophage-tropic CCR5 co-receptor holds particular promise as a target. This receptor is not essential for normal immune function, and individuals homozygous for a 32-bp deletion in this gene are resistant to HIV infection, whereas individuals who are heterozygous for this deletion show delayed progression to AIDS[61,62]. Qin *et al.*[63] used a lentiviral vector to transduce an anti-CCR5 shRNA in human lymphocytes. Downregulation of CCR5 resulted in a modest, but nevertheless significant three- to sevenfold reduction in viral infectivity relative to controls. Despite this downregulation, the CCR5-shRNA-treated cells were still susceptible to infection by the T-tropic virus that uses CXCR4. However, because CXCR4 is essential for the normal function of haematopoietic stem cells[64], targeting this receptor is not a good choice for an anti-HIV therapy, nor is targeting the essential CD4 receptor. So there are drawbacks in targeting cellular HIV cofactors because non-infected cells will inevitably be targeted as well, leading to toxicities that are similar to those observed with the current anti-retroviral drugs. Viral targets will need to be included in any successful strategy using RNAi. These targets should be sequences that are highly conserved throughout the various clades to ensure efficacy against all viral strains.

The delivery of siRNAs or shRNAs to HIV-infected cells is also a challenge. The target cells are primarily T lymphocytes, monocytes and macrophages. As synthetic siRNAs do not persist for long periods in cells, they would have to be delivered repeatedly for years to effectively treat the infection. Systemic delivery of siRNAs to T lymphocytes is probably not feasible owing to the immense number of these cells. Using viral vectors to deliver anti-HIV-encoding shRNA genes is also problematic, and systemic delivery is not yet practicable because the immunogenicity of the vectors themselves precludes performing multiple injections. Therefore the preferred method is to isolate T cells from patients; these T cells are then transduced, expanded and re-infused into the same patients. In a continuing clinical trial, T lymphocytes from HIV-infected individuals are transduced *ex vivo* with a lentiviral vector that encodes an anti-HIV antisense RNA. The transduced cells are subsequently expanded and reinfused into patients[65,66]. This type of therapeutic approach would also be applicable to vectors harbouring genes that encode siRNAs. A different approach is to transduce isolated haematopoietic progenitor or stem cells with vectors harbouring the therapeutic genes. These cells give rise to all the haematopoietic cells capable of being infected by the virus. Haematopoietic stem cells are mobilized from the patient and transduced *ex vivo* before reinfusion (Fig. 3). Two clinical trials in which retroviral vectors expressing ribozymes were transduced into haematopoietic stem cells have demonstrated the feasibility of this approach[67,68]. Because RNAi is more potent than ribozyme or antisense approaches, movement of this technology to a human clinical trial for HIV treatment is expected to take place in the next year or two.

RNAi to treat viral hepatitis

Hepatitis induced by the hepatitis B virus (HBV) and by HCV is a major health problem. At present hundreds of millions of individuals are infected worldwide. There is an effective vaccine against HBV, but this treatment is only useful for the prevention of viral infection and there is no vaccine for HCV. Therefore, hepatitis caused by these two viruses has been an important target for potential RNAi therapy. The first demonstration of RNAi efficacy against a virus *in vivo* involved hydrodynamic co-delivery of an HBV replicon and an expression unit encoding an anti-HBV shRNA in mice[69]. This study demonstrated that a significant knockdown (99%) of the HBV core antigen in liver hepatocytes could be achieved by the shRNA, providing an important proof of principle for future antiviral applications of RNAi in the liver.

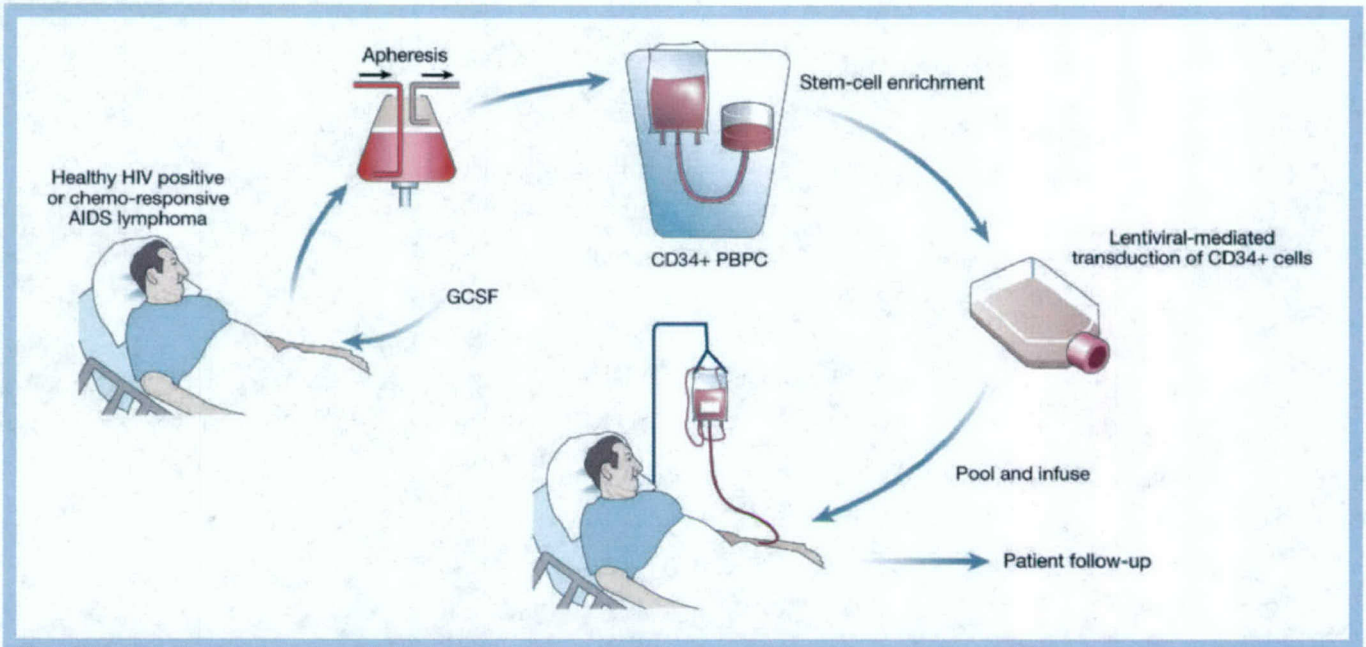


Figure 3 Proposed scheme for the treatment of HIV patients using lentiviral vectors to transduce anti-HIV shRNA genes into the patient's haematopoietic stem cells. Patients are given several injections of granulocyte colony stimulation factor (GCSF), which mobilizes haematopoietic stem cells into the patients' peripheral circulation. Haematopoietic stem cells expressing the CD34 antigen are collected by affinity columns (apheresis) and transduced with a lentiviral vector harbouring the anti-HIV shRNA genes. The cells are then re-infused into patients. Depending on the population, the patient will have been pretreated with no, or with one, or with more than one marrow-chemoablative agent. Following stem-cell engraftment, patients are monitored for a period of several years for HIV loads, CD4+ T-lymphocytes and shRNA gene expression. This overall scheme follows that described by Michienzi *et al.*[68]. PBPC, peripheral blood progenitor cells.

More advanced studies have been carried out for RNAi therapies against HCV, a virus that now infects an estimated 3% of the world's population. HCV is a major cause of chronic liver disease, which can lead to liver cirrhosis and hepatocellular carcinoma. The HCV genome is a positive-strand RNA molecule with a single open reading frame encoding a polyprotein that is processed post-translationally to produce at least ten proteins. The only therapy currently available is a combination of interferon and ribavirin, but response to this therapy is often poor, particularly with certain HCV subtypes.

Subgenomic and full-length HCV replicons that replicate and express HCV proteins in stably transfected human hepatoma-derived Huh-7 cells have been used to study the effects of various antiviral drugs[70–73]. Several groups have now tested the efficacy of siRNA mediated inhibition of replicon function using these systems[74–76]. siRNAs targeting the internal ribosome entry site (IRES) or mRNAs encoding the viral non-structural proteins NS3 and NS5B inhibited HCV replicon function in cell culture[75]. Furthermore, anti-HCV siRNAs depleted Huh-7 cells of persistently replicating HCV replicons[74]. McCaffrey *et al.* performed hydrodynamic tail-vein injections of siRNAs or anti-HCV shRNAs to direct efficient cleavage of HCV sequences in an HCV-luciferase fusion construct in mouse hepatocytes *in vivo*[28].

In another *in vivo* study, siRNAs were used to treat fulminant hepatitis induced by an agonistic Fas-specific antibody in mice. Anti-Fas siRNAs were hydrodynamically injected into the antibody-treated mice: 82% of the mice survived for 10 days of observation whereas all control mice died within 3 days[77]. Importantly, mice already suffering from auto-immune hepatitis also improved after the Fas siRNA treatment. So it may be feasible to use siRNAs to ameliorate the severity of certain diseases by targeting the inflammatory response pathways rather than the infectious agent.

As with HIV therapeutics, delivery of the siRNAs or shRNA vectors is the main challenge for successful treatment of HCV. The method of delivery used in several *in vivo* studies — hydrodynamic intravenous injection — is not feasible for the treatment of human hepatitis. In mice, genetic material can be introduced into hepatocytes using catheters or even localized hydrodynamic procedures[78], but it is yet to be determined whether such procedures can be used to deliver siRNAs in larger mammals.

RNAi and cancer

Many studies have used siRNAs as an experimental tool to dissect the cellular pathways that lead to uncontrolled cell proliferation and to cancer. Moreover, RNAi has been proposed as a potential treatment for cancer[79–81]. Although no clinical trials are yet underway, a precedent might be set by ongoing clinical trials that use antisense reagents. The first systemically delivered antisense oligonucleotide for the treatment of cancer, Genasense (Genta, Inc.), which targets the anti-apoptotic gene *BCL2*, has shown promise in clinical trials for metastatic melanoma when used in combination with conventional chemotherapeutics[82]. However, its use as a US Food and Drug Administration (FDA)-approved drug has recently been put on hold. The potential for using RNAi to treat metastatic cancers will of course depend on finding good cellular targets.

Highly efficient mechanisms for the delivery of siRNA to the relevant cells will also be particularly important for successful treatment of metastatic cancer. Several groups have developed backbone modifications to synthetic siRNAs that provide them with resistance to serum nucleases and should therefore increase the half-life of circulating siRNAs in animal models[83,84]. However, enhancing siRNA stability is not enough unless the siRNAs can penetrate cells and tissue *in vivo* in concentrations sufficient to be therapeutically functional. As siRNAs are double-stranded molecules, delivery and cellular uptake is more of a challenge than for single-stranded antisense agents, which bind to serum proteins and are taken up by cells and tissues *in vivo*[85]. There are a few reports of functional RNAi being obtained by systemic delivery of liposome-encapsulated siRNAs, but the use of cationic or anionic lipids for *in vivo* delivery of antisense agents has never reached a clinical trial. Therefore, we still need to understand better which backbone modifications might be useful for enhancing cellular and tissue uptake of naked RNAs, or we need to develop alternative carriers for systemic delivery of siRNAs — a feat that will be essential in treating metastatic cancers.

Using RNAi to target genes expressing oncogenic fusion proteins, such as the Bcr-Abl oncoprotein p210 that is characteristic of chronic myelogenous leukaemia (CML), has provided excellent proof of principle for RNAi as an anti-cancer therapeutic agent. For CML, the main treatment options have been chemotherapy, allogeneic bone-marrow transplant and most recently, the use of a small molecule, the tyrosine kinase inhibitor, imatinib. Despite initial excitement about the potential of imatinib, a growing number of patients have developed resistance to it[86–89], necessitating alternative forms of therapy. Bcr-Abl p210 has been selectively downregulated by both synthetic siRNAs and lentiviral-vector-transduced shRNAs in cell lines[90–92]. Importantly, the downregulation is selective for only the p210 oncoprotein and its mRNA, which results in inhibition of cell proliferation as a direct consequence of RNAi. Haematologic malignancies are often treated by bone-marrow transplantation. Therefore, a possible therapeutic application would be to transduce haematopoietic stem cells with vectors harbouring a gene that targets the mRNA encoding the oncogenic p210 protein, thereby protecting patients from relapse caused by proliferation of latent leukaemic stem cells. Again, delivery is the key issue; 100% transduction of the stem cells reinfused in a bone-marrow transplant setting will be required to make this therapeutically effective. The improvements in viral vector titres and transduction efficiencies may make this possible.

RNAi for genetic diseases

A promising lead towards using RNAi for the treatment of genetic diseases has been provided by preliminary studies that demonstrate how SNPs in mutant allele transcripts can be used as selective targets for RNAi[93,94]. Finding an siRNA that is highly selective for a particular SNP is a challenge, but has been accomplished by systematic analyses of siRNAs in which the polymorphic nucleotide is complementary to the mid-region of the siRNA. In certain examples, the siRNAs direct selective degradation of only the mutant transcripts, leaving the wild-type transcripts intact despite only a single mismatch[93,94]. Another example of siRNAs targeting an SNP was recently reported in studies of amyotrophic lateral sclerosis (ALS) caused by mutations in the Cu, Zn superoxide dismutase (*SOD1*) gene[95]. Because the wild-type SOD1 performs important functions, it is important to selectively eliminate expression of only the mutant allelic transcript. Many SOD1 mutations are single-nucleotide changes. Ding *et al.*[95] achieved selective degradation of a mutant *SOD1* allele, thereby providing a potential therapeutic application for the treatment of ALS.

Disease-causing polyglutamine proteins encoded by CAG-repeat-containing transcripts are found in several neurological diseases such as Huntington's disease. These proteins are especially challenging targets for RNAi because CAG repeats are common to many normal transcripts as well, and the repeats themselves cannot be selectively targeted by siRNAs. But with the recent finding that delivery of siRNAs and viral vectors expressing siRNAs to diseased regions of the brain is technically feasible[96], coupled with selective targeting of SNPs in the mutant transcripts, the promise of clinical use of RNAi for the treatment of degenerative, neurological diseases should be realised.

Challenges for RNAi as a therapy

Two key challenges in developing RNAi as a therapy are avoiding off-target effects and ensuring efficient delivery. One potential risk for side effects emerges from the feature that distinguishes RNAi from other antisense technologies — the use of cellular machinery for directing sequence-specific silencing. This machinery has specific purposes, such as miRNA-mediated gene regulation[97,98]. Using siRNAs to target specific cellular or viral transcripts in essence hijacks the endogenous RNAi machinery, and we know little about the potential for saturating the RNAi pathway in primary cells, although saturation of RISC is demonstrable in cultured cells[99]. So endogenous RNAi pathways could potentially be affected by siRNAs. It will be important to pay close attention to basic research studies on off-target effects of siRNAs and on the design of effective siRNAs[22,27,100]. A better understanding of the mechanisms that lead to nonspecific effects of short dsRNAs is essential before the use of siRNAs or shRNAs can be tested in patient trials.

The issue of delivery has restricted the antisense field for almost two decades. It is feasible to infuse backbone-modified oligonucleotides *in vivo*, but achieving intracellular delivery at therapeutically effective concentrations is a major challenge. Targeted delivery to specific cell or tissue types is still not a practical reality for oligonucleotide-based therapeutics. The alternative approach is viral-vector-mediated delivery of therapeutic shRNA genes. Because this type of delivery results in gene therapy, there are several associated safety concerns, and systemic delivery of viral vectors is still a major hurdle. Nevertheless, the potency and potential general therapeutic utility of RNAi is prompting renewed vigour into delivery-related research. It remains to be determined whether backbone-modified, nuclease-resistant siRNAs will move to the clinic more quickly than synthetic deoxyoligonucleotides.

Perspective

In a remarkably short time since its discovery in model organisms, the RNAi pathway has emerged as a powerful tool for the study of gene function in mammals. As our understanding of the under-lying biology and biochemistry of this conserved gene-regulatory mechanism improves, so does our ability to exploit RNAi as an experimental tool. With the use of RNAi in whole animals increasing, we anticipate growing enthusiasm for the use of RNAi triggers in therapy. Despite considerable hurdles to overcome, it seems likely that RNAi will find a place alongside more conventional approaches in the treatment of diseases, although it is unclear how long we will have to wait to witness the first RNAi-based drug. The big question is whether RNAi can revolutionize the treatment of human disease in the same way that it has revolutionized basic research into gene function. □

doi:10.1038/nature02870

1. Fire, A., Albertson, D., Harrison, S. W. & Moerman, D. G. Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development* **113,** 503–514 (1991).
2. Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81,** 611–620 (1995).
3. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811 (1998).
4. Hannon, G. J. RNA interference. *Nature* **418,** 244-251 (2002).
5. Williams, B. R. Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. *Biochem. Soc. Trans.* **25,** 509–513 (1997).
6. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366 (2001).
7. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293,** 1146–1150 (2001).
8. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494–498 (2001).
9. Hutvágner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293,** 834–838 (2001).
10. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106,** 23–34 (2001).
11. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15,** 2654–2659 (2001).
12. Paddison, P. J., Caudy, A. A., Sachidanandam, R. & Hannon, G. J. Short hairpin activated gene silencing in mammalian cells. *Methods Mol. Biol.* **265,** 85–100 (2004).
13. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115,** 209–216 (2003).
14. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115,** 199–208 (2003).
15. Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nature Biotechnol.* **22,** 326–330 (2004).
16. Silva, J. M., Sachidanandam, R. & Hannon, G. J. Free energy lights the path toward more effective RNAi. *Nature Genet.* **35,** 303–305 (2003).
17. Aza-Blanc, P. *et al.* Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12,** 627–637 (2003).
18. Pebernard, S. & Iggo, R. D. Determinants of interferon-stimulated gene induction by RNAi vectors. *Differentiation* **72,** 103–111 (2004).
19. Persengiev, S. P., Zhu, X. & Green, M. R. Nonspecific, concentration-dependent stimulation and repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA* **10,** 12–18 (2004).
20. Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H. & Williams, B. R. Activation of the interferon system by short-interfering RNAs. *Nature Cell Biol.* **5,** 834–839 (2003).
21. Kim, D. H. *et al.* Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. *Nature Biotechnol.* **22,** 321–325 (2004).
22. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21,** 635–637 (2003).
23. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18,** 504–511 (2004).

24. Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4,** R42 (2003).
25. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115,** 787–798 (2003).
26. Enright, A. J. *et al.* MicroRNA targets in *Drosophila. Genome Biol.* **5,** R1 (2003).
27. Scacheri, P. C. *et al.* Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **101,** 1892–1897 (2004).
28. McCaffrey, A. P. *et al.* RNA interference in adult mice. *Nature* **418,** 38–39 (2002).
29. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro. Proc. Natl Acad. Sci. USA* **101,** 16–22 (2004).
30. Layzer, J. M. *et al. In vivo* activity of nuclease-resistant siRNAs. *RNA* **10,** 766–771 (2004).
31. Lewis, D. L., Hagstrom, J. E., Loomis, A. G., Wolff, J. A. & Herweijer, H. Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nature Genet.* **32,** 107–108 (2002).
32. Sorensen, D. R., Leirdal, M. & Sioud, M. Gene silencing by systemic delivery of synthetic siRNAs in adult mice. *J. Mol. Biol.* **327,** 761–766 (2003).
33. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2,** 243–247 (2002).
34. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33,** 401–406 (2003).
35. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo. Nature Genet.* **33,** 396–400 (2003).
36. Carmell, M. A., Zhang, L., Conklin, D. S., Hannon, G. J. & Rosenquist, T. A. Germline transmission of RNAi in mice. *Nature Struct. Biol.* **10,** 91–92 (2003).
37. Tiscornia, G., Singer, O., Ikawa, M. & Verma, I. M. A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. *Proc. Natl Acad. Sci. USA* **100,** 1844–1848 (2003).
38. Hasuwa, H., Kaseda, K., Einarsdottir, T. & Okabe, M. Small interfering RNA and gene silencing in transgenic mice and rats. *FEBS Lett.* **532,** 227–230 (2002).
39. Kunath, T. *et al.* Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nature Biotechnol.* **21,** 559–561 (2003).
40. Hsieh, A. C. *et al.* A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32,** 893–901 (2004).
41. Brummelkamp, T. R., Nijman, S. M., Dirac, A. M. & Bernards, R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-κB. *Nature* **424,** 797–801 (2003).
42. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428,** 431–437 (2004).
43. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428,** 427–431 (2004).
44. Zheng, L. *et al.* An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA* **101,** 135–140 (2004).
45. Luo, B., Heard, A. D. & Lodish, H. F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc. Natl Acad. Sci. USA* **101,** 5494–5499 (2004).
46. Shirane, D. *et al.* Enzymatic production of RNAi libraries from cDNAs. *Nature Genet.* **36,** 190-196 (2004).
47. Sen, G., Wehrman, T. S., Myers, J. W. & Blau, H. M. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nature Genet.* **36,** 183–189 (2004).
48. Silva, J. M., Mizuno, H., Brady, A., Lucito, R. & Hannon, G. J. RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. *Proc. Natl Acad. Sci. USA* **101,** 6548–6552 (2004).
49. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110 (2001).
50. Jacque, J. M., Triques, K. & Stevenson, M. Modulation of HIV-1 replication by RNA interference. *Nature* **418,** 435–438 (2002).
51. Lee, N. S. *et al.* Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nature Biotechnol.* **20,** 500–505 (2002).
52. Coburn, G. A. & Cullen, B. R. Potent and specific inhibition of human immunodeficiency virus type-1 replication by RNA interference. *J. Virol.* **76,** 9225–9231 (2002).
53. Surabhi, R. M. & Gaynor, R. B. RNA interference directed against viral and cellular targets inhibits human immunodeficiency virus type-1 replication. *J. Virol.* **76,** 12963–12973 (2002).
54. Novina, C. D. *et al.* siRNA-directed inhibition of HIV-1 infection. *Nature Med.* **8,** 681–686 (2002).
55. Park, W. S. *et al.* Prevention of HIV-1 infection in human peripheral blood mononuclear cells by specific RNA interference. *Nucleic Acids Res.* **30,** 4830–4835 (2002).
56. Boden, D., Pusch, O., Lee, F., Tucker, L. & Ramratnam, B. Human immunodeficiency virus type-1 escape from RNA interference. *J. Virol.* **77,** 11531–11535 (2003).
57. Martinez, M. A. *et al.* Suppression of chemokine receptor expression by RNA interference allows for inhibition of HIV-1 replication. *AIDS* **16,** 2385–2390 (2002).
58. Capodici, J., Kariko, K. & Weissman, D. Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. *J. Immunol.* **169,** 5196–5201 (2002).
59. Banerjea, A. *et al.* Inhibition of HIV-1 by lentiviral vector-transduced siRNAs in T lymphocytes differentiated in SCID-hu mice and CD34+ progenitor cell-derived macrophages. *Mol. Ther.* **8,** 62–71 (2003).
60. Li, M. J. *et al.* Inhibition of HIV-1 infection by lentiviral vectors expressing Pol III-promoted anti-HIV RNAs. *Mol. Ther.* **8,** 196–206 (2003).
61. Eugen-Olsen, J. *et al.* Heterozygosity for a deletion in the *CKR-5* gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. *AIDS* **11,** 305–310 (1997).
62. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. *Nature* **382,** 722–725 (1996).
63. Qin, X. F., An, D. S., Chen, I. S. & Baltimore, D. Inhibiting HIV-1 infection in human T cells by lentiviral-mediated delivery of small interfering RNA against CCR5. *Proc. Natl Acad. Sci. USA* **100,** 183–188 (2003).
64. Dell'Agnola, C. *et al. In vitro* and *in vivo* hematopoietic potential of human stem cells residing in muscle tissue. *Exp. Hematol.* **30,** 905–914 (2002).
65. Dropulic, B. Lentivirus in the clinic. *Mol. Ther.* **4,** 511–512 (2001).
66. Davis, B. M., Humeau, L. & Dropulic, B. *In vivo* selection for human and murine hematopoietic cells transduced with a therapeutic MGMT lentiviral vector that inhibits HIV replication. *Mol. Ther.* **9,** 160–172 (2004).
67. Amado, R. G. *et al.* Anti-human immunodeficiency virus hematopoietic progenitor cell-delivered ribozyme in a phase I study: myeloid and lymphoid reconstitution in human immunodeficiency virus type-1-infected patients. *Hum. Gene Ther.* **15,** 251–262 (2004).
68. Michienzi, A. *et al.* RNA-mediated inhibition of HIV in a gene therapy setting. *Ann. NY Acad. Sci.* **1002,** 63–71 (2003).
69. McCaffrey, A. P. *et al.* Inhibition of hepatitis B virus in mice by RNA interference. *Nature Biotechnol.* **6,** 639–644 (2003).
70. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290,** 1972–1974 (2000).
71. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285,** 110–113 (1999).
72. Ikeda, M., Yi, M., Li, K. & Lemon, S. M. Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. *J. Virol.* **76,** 2997–3006 (2002).
73. Pietschmann, T., Lohmann, V., Rutter, G., Kurpanek, K. & Bartenschlager, R. Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J. Virol.* **75,** 1252–1264 (2001).
74. Randall, G., Grakoui, A. & Rice, C. M. Clearance of replicating hepatitis C virus replicon RNAs in cell culture by small interfering RNAs. *Proc. Natl Acad. Sci. USA* **100,** 235–240 (2003).
75. Wilson, J. A. *et al.* RNA interference blocks gene expression and RNA synthesis from hepatitis C replicons propagated in human liver cells. *Proc. Natl Acad. Sci. USA* **100,** 2783–2788 (2003).
76. Kapadia, S. B., Brideau-Andersen, A. & Chisari, F. V. Interference of hepatitis C virus RNA replication by short interfering RNAs. *Proc. Natl Acad. Sci. USA* **100,** 2014–2018 (2003).
77. Song, E. *et al.* RNA interference targeting Fas protects mice from fulminant hepatitis. *Nature Med.* **9,** 347–351 (2003).
78. Eastman, S. J. *et al.* Development of catheter-based procedures for transducing the isolated rabbit liver with plasmid DNA. *Hum. Gene Ther.* **13,** 2065–2077 (2002).
79. Kittler, R. & Buchholz, F. RNA interference: gene silencing in the fast lane. *Semin. Cancer Biol.* **13,** 259–265 (2003).
80. Wall, N. R. & Shi, Y. Small RNA: can RNA interference be exploited for therapy? *Lancet* **362,** 1401–1403 (2003).
81. Lu, P. Y., Xie, F. Y. & Woodle, M. C. siRNA-mediated antitumorigenesis for drug target validation and therapeutics. *Curr. Opin. Mol. Ther.* **5,** 225–234 (2003).
82. Buchele, T. Proapoptotic therapy with oblimersen (bcl-2 antisense oligonucleotide)—review of preclinical and clinical results. *Onkologie* **26** (Suppl. 7), 60–69 (2003).
83. Chiu, Y. L. & Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **9,** 1034–1048 (2003).
84. Czauderna, F. *et al.* Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.* **31,** 2705–2716 (2003).
85. Wang, L., Prakash, R. K., Stein, C. A., Koehn, R. K. & Ruffner, D. E. Progress in the delivery of therapeutic oligonucleotides: organ/cellular distribution and targeted delivery of oligonucleotides *in vivo. Antisense Nucleic Acid Drug Dev.* **13,** 169–189 (2003).
86. Holtz, M. S. & Bhatia, R. Effect of imatinib mesylate on chronic myelogenous leukemia hematopoietic progenitor cells. *Leuk. Lymphoma* **45,** 237–245 (2004).
87. Tauchi, T. & Ohyashiki, K. Molecular mechanisms of resistance of leukemia to imatinib mesylate. *Leuk. Res.* **28** (Suppl. 1), 39–45 (2004).
88. Cowan-Jacob, S. W. *et al.* Imatinib (STI571) resistance in chronic myelogenous leukemia: molecular basis of the underlying mechanisms and potential strategies for treatment. *Mini Rev. Med. Chem.* **4,** 285–299 (2004).
89. Marcucci, G., Perrotti, D. & Caligiuri, M. A. Understanding the molecular basis of imatinib mesylate therapy in chronic myelogenous leukemia and the related mechanisms of resistance. Commentary. *Clin. Cancer Res.*, **9:** 1333–1337, 2003. *Clin. Cancer Res.* **9,** 1248–1252 (2003).
90. Li, M. J., McMahon, R., Snyder, D. S., Yee, J. K. & Rossi, J. J. Specific killing of Ph+ chronic myeloid leukemia cells by a lentiviral vector-delivered anti-bcr/abl small hairpin RNA. *Oligonucleotides* **13,** 401–409 (2003).
91. Wohlbold, L. *et al.* Inhibition of *bcr-abl* gene expression by small interfering RNA sensitizes for imatinib mesylate (STI571). *Blood* **102,** 2236–2239 (2003).
92. Scherr, M. *et al.* Specific inhibition of *bcr-abl* gene expression by small interfering RNA. *Blood* **101,** 1566–1569 (2003).
93. Miller, V. M., Gouvion, C. M., Davidson, B. L. & Paulson, H. L. Targeting Alzheimer's disease genes with RNA interference: an efficient strategy for silencing mutant alleles. *Nucleic Acids Res.* **32,** 661–668 (2004).
94. Miller, V. M. *et al.* Allele-specific silencing of dominant disease genes. *Proc. Natl Acad. Sci. USA* **100,** 7195–7200 (2003).
95. Ding, H. *et al.* Selective silencing by RNAi of a dominant allele that causes amyotrophic lateral sclerosis. *Aging Cell* **2,** 209–217 (2003).
96. Davidson, B. L. & Paulson, H. L. Molecular medicine for the brain: silencing of disease genes with RNA interference. *Lancet Neurol.* **3,** 145–149 (2004).
97. Pasquinelli, A. E. MicroRNAs: deviants no longer. *Trends Genet.* **18,** 171–173 (2002).
98. Moss, E. G. MicroRNAs: hidden in the genome. *Curr. Biol.* **12,** R138–R140 (2002).
99. Hutvágner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol.* **2,** E98 (2004).
100. Saxena, S., Jonsson, Z. O. & Dutta, A. Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J. Biol. Chem.* **278,** 44312–44319 (2003).

Acknowledgements We thank members of the RNAi community for making this an exciting field of research. G.J.H. thanks F. Rivas for comments on the manuscript and J. Duffy for help with the figures. J.J.R. thanks L. Scherer for suggestions on therapeutic applications of RNAi. G.J.H. is supported by an Innovator award from the US Army Breast Cancer Research Program and by grants from the NIH. J.J.R. is supported by grants from the NIH, NIAID and NHBLI.

Competing interests statement The authors declare that they have no competing financial interests.

nature biotechnology

Synthetic shRNAs as potent RNAi triggers

Despina Siolas[1,2], Cara Lerner[3], Julja Burchard[3], Wei Ge[3], Peter S Linsley[3], Patrick J Paddison[2], Gregory J Hannon[2] & Michele A Cleary[3]

Designing potent silencing triggers is key to the successful application of RNA interference (RNAi) in mammals. Recent studies suggest that the assembly of RNAi effector complexes is coupled to Dicer cleavage. Here we examine whether transfection of optimized Dicer substrates results in an improved RNAi response. Dicer cleavage of chemically synthesized short hairpin RNAs (shRNAs) with 29-base-pair stems and 2-nucleotide 3′ overhangs produced predictable homogeneous small RNAs comprising the 22 bases at the 3′ end of the stem. Consequently, direct comparisons of synthetic small interfering RNAs and shRNAs that yield the same small RNA became possible. We found synthetic 29-mer shRNAs to be more potent inducers of RNAi than small interfering RNAs. Maximal inhibition of target genes was achieved at lower concentrations and silencing at 24 h was often greater. These studies provide the basis for an improved approach to triggering experimental silencing via the RNAi pathway.

Many eukaryotic organisms respond to double-stranded RNA (dsRNA) by activating a sequence-specific silencing pathway RNAi. RNAi is initiated when an RNase III–family nuclease, Dicer, processes dsRNAs into ~22-nucleotide (nt) fragments known as small interfering (siRNAs)[1–3]. These small RNAs are used as guides for selection and cleavage of complementary mRNAs through their incorporation into the RNAi effector complex (RISC)[1,2,4], whose catalytic subunit, Argonaute 2, has recently been identified[5,6]. These mechanistic insights have led to approaches for experimentally programming the RNAi machinery in mammalian cells by directly transfecting chemically synthesized siRNA duplexes of ~21 nt, consisting of 19 paired bases with 2-nt 3′ overhangs, to produce a transient silencing response[7].

In many organisms, the RNAi machinery also serves as an effector for endogenous, noncoding RNAs known as microRNAs (miRNAs)[8]. miRNAs are initially generated as long primary transcripts (pri-miRNA) which are cleaved in the nucleus by another RNase III–family nuclease, Drosha[9]. The liberated pre-miRNAs are exported to the cytoplasm, where Dicer performs a second cleavage to produce small RNAs that are loaded into RISC[10–12]. In the case of miRNAs, the cleavage sites are specific, and most often a single, discrete sequence is liberated from the precursor[8]. These discoveries prompted the development of a second approach for triggering RNAi in mammalian cells using DNA vectors encoding shRNAs, modeled roughly after endogenous microRNAs[13–15].

Remarkably, for both miRNAs and siRNAs, the two strands of the processed dsRNA are treated unequally. In a variety of organisms, cloning has overwhelmingly yielded one strand for each miRNA[8]. A potential explanation for this outcome came from biochemical studies of siRNAs in *Drosophila melanogaster* that suggested that relative thermodynamic instability at the 5′ end of a strand of a Dicer product favors its loading into RISC[16]. This is in accord with analysis of predicted Dicer cleavage products of endogenous miRNAs[17] and studies of the efficacy of large numbers of siRNAs, which indicate that greater suppression occurs if the antisense strand (relative to the target mRNA) has an unstable 5′ end[17]. Recent reports have suggested that this loading might occur in a complex and might be coordinated with Dicer cleavage[18–20]. These mechanistic insights suggest that Dicer substrates might be more efficiently incorporated into RISC than siRNAs. To compare the efficiency of silencing triggers predicted to produce equivalent RISC enzymes, we sought to understand how Dicer processes shRNAs.

We began by producing ~70 chemically synthesized shRNAs, targeting various endogenous genes and reporters. We focused on a detailed analysis of one set of four shRNAs that target firefly luciferase (**Fig. 1a**). The individual species differed in two ways. First, the stems of the shRNAs were either 19 or 29 base pairs (bp) long; these sizes reflect the two stem sizes most commonly used for vector-expressed shRNAs. Second, each shRNA either did or did not contain a 2-nt 3′ overhang, identical to that produced by the processing of pri-miRNAs by Drosha. Each species was end labeled by enzymatic phosphorylation and incubated with recombinant human Dicer. The 29-mer shRNA with the 3′ overhang was converted almost quantitatively into a 22-nt product by Dicer (**Fig. 1b**). In contrast, the 29-mer shRNA without the overhang generated very little discrete 22-nt labeled product, despite a Dicer-dependent depletion of the starting material. Neither 19-mer shRNA was cleaved to a detectable level by the Dicer enzyme. This result was not due to the lack of double-stranded structure in the 19-mer shRNAs, as all shRNA substrates were efficiently cleaved by bacterial RNase III (**Supplementary Fig. 1** online). Rather, these results suggest that the shRNAs with a 3′ overhang produced predominantly one specific and unique small RNA product, whereas a blunt-ended hairpin was processed into a range of products. This hypothesis was consistent with parallel analysis

[1]Program in Genetics, Stony Brook University, Stony Brook, New York 11794, USA. [2]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [3]Rosetta Inpharmatics, LLC, a wholly owned subsidiary of Merck and Co., Inc., 401 Terry North, Seattle, Washington 98109, USA. Correspondence should be addressed to G.J.H. (hannon@cshl.edu) and M.A.C. (michele_cleary@merck.com).

Published online 26 December 2004; doi:10.1038/nbt1052

of identical shRNA substrates that were produced by *in vitro* transcription with T7 polymerase and uniformly labeled (**Fig. 1c**). Uniformly labeled 29-mer shRNAs both with and without overhangs produced cleavage products, with the latter being less abundant. Additionally, shRNAs with overhangs yielded products of two discrete sizes (21 and 22 nt). Considered together, our results suggest that Dicer requires a minimum stem length for efficient cleavage. Furthermore, they are consistent with the hypothesis that the presence of a correct 3′ overhang enhances the efficiency and specificity of cleavage, directing Dicer to cut ~22 nt from the end of the substrate.

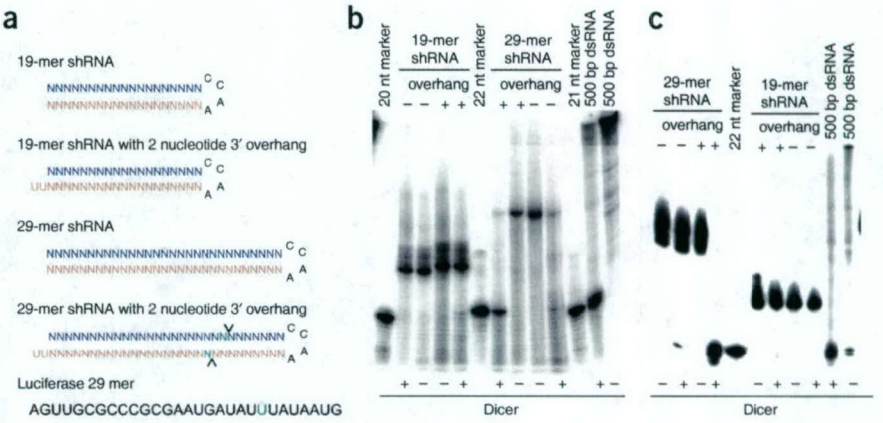


Figure 1 *In vitro* processing of 29-bp shRNAs by Dicer generates a predominant siRNA from the end of each short hairpin. (**a**) The set of shRNAs containing 19- or 29-bp stems and either with or without a 2-nt 3′ overhang is depicted schematically. For reference, the 29-nt sequence from luciferase (top, blue) strand is given. The presumed cleavage sites (as predicted by analysis of Dicer processing products) are indicated in green and by the arrows. (**b**) *In vitro* Dicer processing of shRNAs. 5′ end labeled substrates as depicted in **a** were incubated either in the presence or absence of recombinant human Dicer. Processing of a 500-bp blunt-ended dsRNA is shown for comparison. Markers are end labeled single-stranded synthetic RNA oligonucleotides. (**c**) Uniformly labeled shRNAs with structures as indicated in **a** were processed by Dicer to produce a small RNA product. Results of processing a 500-bp blunt-ended dsRNA are shown for comparison.

A number of previous studies have suggested that Dicer might function as an end-recognizing endonuclease without positing a role for the 3′ overhang. Blocking the ends of dsRNAs using either fold-back structures or chimeric RNA-DNA hybrids attenuated, but did not abolish, the ability of human Dicer to generate siRNAs[21]. One group suggested that Dicer cleaved ~22 nt from the blunt end of an extended pre-miRNA, designed in part to mimic a pri-miRNA[22]. Structural analysis of the Argonaute 2 PAZ domain suggested that it engages very short (~2- to 3-nt) stretches of the 3′ ends of single-stranded RNAs[23–26]. This led another group of researchers to propose a model in which the 3′ overhangs of pre-miRNAs, generated by Drosha cleavage, serve as an important recognition and specificity determinant for subsequent processing by Dicer[25]. The results presented here are consistent with this model and suggest further that the 3′ overhang aids in determining the specificity of cleavage, directing processing to a site 22 nt from the 3′ end of the substrate. These findings are in full accord with a recently published model for Dicer action[27].

To validate our biochemical analysis, we also mapped the position of Dicer cleavage *in vivo* using primer extension. Precursors were transfected into cells, and the processed form of each was isolated by coimmunoprecipitation with the coexpressed Myc-tagged human Argonaute proteins Ago1 and Ago2. The 29-mer shRNA with an overhang gave rise to a relatively discrete product of 20 nt as predicted for a cleavage 22 nt from the 3′ end of the substrate. Primer extension suggested identical cleavage specificities when shRNAs were exposed to Dicer either *in vitro* or in living cells (**Fig. 2a**). Control experiments using a luciferase 29-mer shRNA alone (without Myc-tagged Ago1 or Ago2 expression) or cells transfected with Myc-tagged Ago1 or Ago2 alone (no shRNA) did not yield extension products (**Fig. 2b**).

Although the inability of Dicer to effectively cleave shRNAs with 19-bp stems may seem at odds with the effective use of such structures for triggering RNAi using vector-based expression, there is presently no evidence that these RNAs require Dicer for their action. Indeed, our results using RNAi to deplete Dicer from cells suggest a strong dependence on Dicer for shRNAs with 29-bp stems, but little dependence for shRNAs with 19-bp stems (not shown). However, 19-mer shRNA do enter RISC. Human 293 cells that constitutively express Ago1 were transfected with siRNAs, 29-mer shRNAs or 19-mer shRNAs. RISC was recovered by immunoprecipitation and associated RNAs were examined by northern blotting. (**Supplementary Fig. 2** online) The 29-mer shRNA with an overhang and the 22-mer siRNA both entered RISC, producing 22-nt small RNAs. The 19-mer shRNA

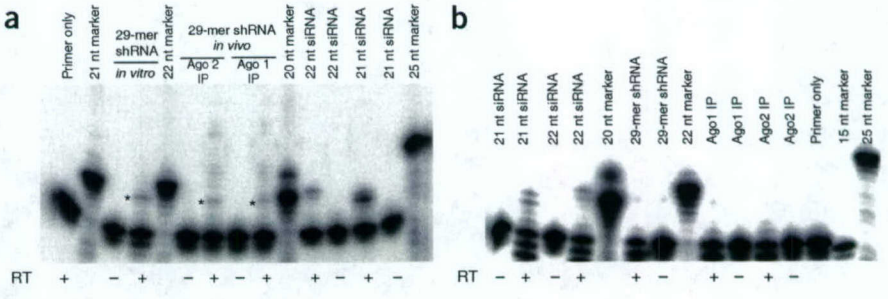


Figure 2 Primer extension analysis shows that similar small RNAs are generated by Dicer processing *in vitro* or *in vivo*. (**a**) Primer extension was used to analyze products from processing of overhang-containing 29-mer shRNAs *in vivo*. Total RNAs were extended with a specific primer that yields a 20-base product if cleavage occurs 22 bases from the 3′ end of the overhang-containing RNA (see **Fig. 1a**). For comparison, extensions of *in vitro* processed material are also shown. Lanes labeled siRNA are extensions of synthetic RNAs corresponding to predicted siRNAs that would be released by cleavage 21 or 22 nt from the 3′ end of the overhang-containing precursor. Observation of extension products depends entirely on the inclusion of reverse transcriptase (RT). The * indicates the specific extension product. Markers are phosphorylated, synthetic DNA oligonucleotides. (**b**) Total RNA from control transfections, which lacked a coexpressed tagged Ago protein, making it impossible to recover small RNAs in the immunoprecipitates, did not show a primer extension product. The same primer was used for all extensions and is compatible with all RNAs. Controls labeled Ago1 or Ago2 lacked co-transfected target RNAs.

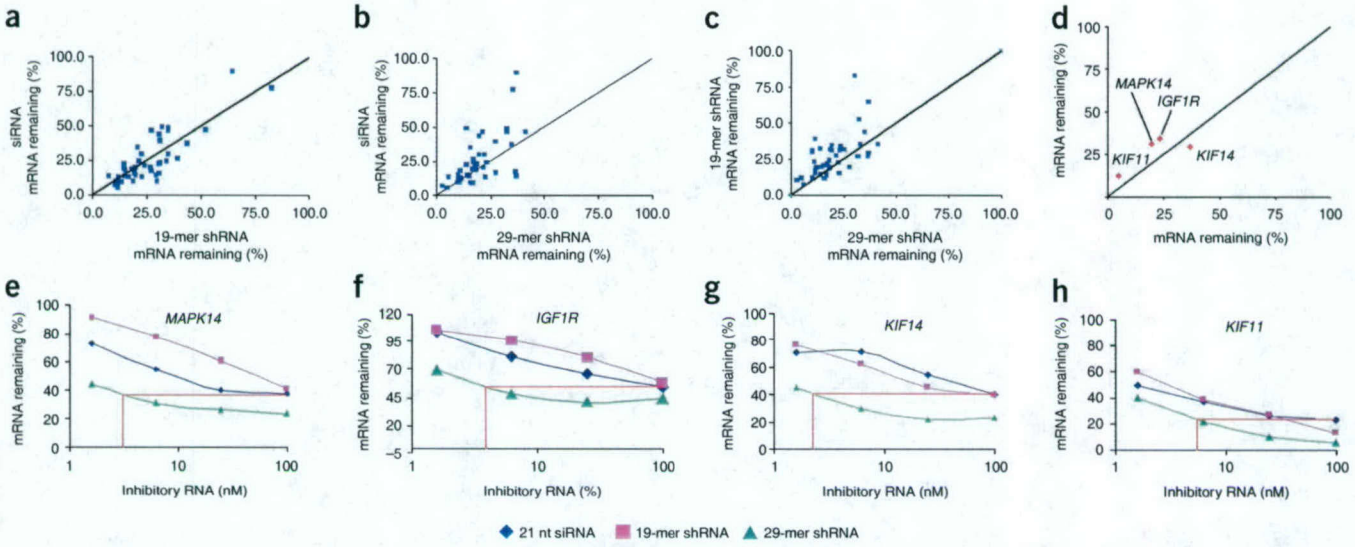


Figure 3 Gene suppression by shRNAs is comparable to or more effective than that achieved by siRNAs targeting the same sequences. (**a–c**) mRNA suppression by 43 siRNAs targeting six different genes was compared with suppression by 19-mer or 29-mer shRNAs derived from the same target sequences. 19-mer and 29-mer shRNAs were also directly compared. All RNAs were transfected at a final concentration of 100 nM. Values indicated on the *x* and *y* axes reflect the percentage of mRNA remaining after 24 h in HeLa cells transfected with RNA as compared with cells treated with transfection reagent alone. (**d–h**) Four representative sets of siRNA and 19-mer and 29-mer shRNAs were used in dose-response analysis to compare the potency of representative RNAi triggers targeting four genes. Comparisons of relative suppression (19-mer versus 29-mer) at the maximal dose are shown for reference in **d**. Titration curves were also performed reporting the percentage of target mRNA remaining (*y* axis) from transfections at 1.56, 6.25, 25 and 100 nM final concentrations of RNA (*x* axis). Percentage of RNA remaining was determined by semiquantitative RT-PCR. Gene targets were *MAPK14*, *KIF11*, *IGF1R* and *KIF14*. (Sequences used were MAPK14-4, KIF11-6, IGF1R-1, KIF14-1 as in **Supplementary Table 1**.) Blue diamonds, 21-mer siRNAs; pink squares, 19-mer shRNAs; green triangles, 29-mer shRNAs. Red lines indicate the concentration of 29-mer shRNA that gives the level of inhibition achieved by 100 nM siRNA.

also entered RISC but produced two distinct small RNAs of 21 and 23 nt. Although we do not understand the mechanistic basis for this observation, it may reflect Dicer-independent cleavage of the 19-mer shRNA in the loop by a single-strand specific ribonuclease.

Because we could predict which single, specific 22-nt sequence would be incorporated into RISC from a given shRNA, we could directly compare the silencing efficiency of shRNAs and siRNAs. Toward this goal, we selected 43 sequences targeting a total of 5 genes (3–9 sequences per gene). For each sequence, we synthesized a 21-mer siRNA (19-bp stem) and shRNAs with 19- or 29-bp stems that were predicted to give Dicer products that either were identical to their corresponding siRNAs or differed by the addition of one 3′ nucleotide homologous to the target. Importantly, each was predicted to give precisely the same 5′ end after cleavage of a 22-mer RNA from the shRNA (**Supplementary Fig. 3** online). Sequences for siRNAs are provided in **Supplementary Table 1** online. Each RNA species was transfected into HeLa cells at a relatively high concentration (100 nM). The level of suppression was determined by semiquantitative RT-PCR of RNA from HeLa cells 24 h after transfection and the performance of each shRNA was compared with the performance of the corresponding siRNA. Studies assessing siRNAs and 19-mer shRNAs showed that there was little difference in silencing at 24 h with these species (**Fig. 3a**). A comparison of siRNAs with shRNAs having 29-bp stems gave a different result. Clustering of the data points above the diagonal indicated consistently better inhibition with the 29-mer shRNAs (**Fig. 3b**). As predicted from the aforementioned results, direct comparisons of shRNAs with 19- and 29-bp stems indicated a greater overall effect with the latter structure (**Fig. 3c**).

The generally better inhibition with 29-mer shRNAs at a high dose led us to investigate the potency of these silencing triggers as compared with siRNAs. Seventeen complete sets comprising an siRNA, a 19-mer shRNA and a 29-mer shRNA were examined for suppression in titration experiments. In no case did the 19-mer shRNAs perform better than the corresponding siRNAs. In contrast, 29-mer shRNAs exceeded the performance of siRNAs in the majority of cases. In most cases, the 29-mer shRNAs showed greater inhibition at the maximal dose; however, even when this inhibition at the maximal dose did not differ much from the siRNA or 19-mer shRNA, the efficacy of the 29-mer at lower concentrations was substantially better. The dose-response experiments for four representative sets of RNAs are shown in **Figure 3d–h**.

Consistent with our results for most of the RNA sets tested, in the case of *MAPK14*, *KIF14* and *KIF11*, the maximal level of suppression for the 29-mer shRNA was approximately twofold greater than the maximal level of suppression for the corresponding siRNA (**Fig. 3e–h**). More importantly, in some cases, the amount of RNA required to achieve maximal inhibition was up to 20-fold lower with 29-mer shRNA than with a similar 21-mer siRNA. This greater potency for 29-mer shRNA as compared to the other two RNA species may reflect the entry of these RNAs into the RNAi pathway as natural intermediates and may explain their greater efficacy when delivered from vectors[14].

Microarray analysis has shown downregulation of many nontargeted transcripts after transfection of siRNAs into HeLa cells[28]. Notably, these gene expression signatures differed between different siRNAs targeting the same gene. Many of the 'off-target' transcripts contained sites of partial identity to the individual siRNA, possibly explaining the source of the effects. To examine potential off-target effects of synthetic shRNAs, we compared shRNA signatures with those of siRNAs derived from the same target sequence.

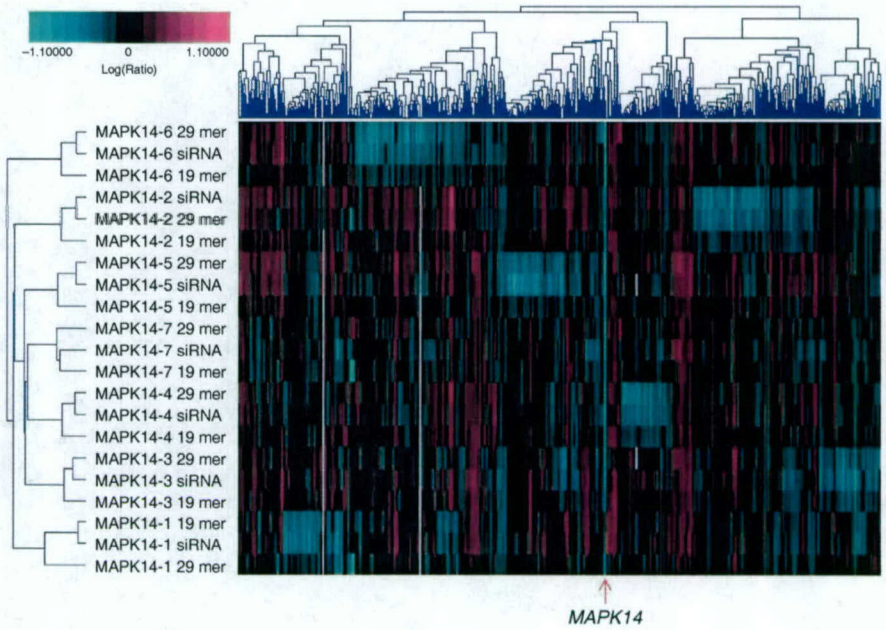


Figure 4 Microarray profiling shows that gene expression profiles of 29-mer shRNAs and the corresponding siRNAs are more similar than expression profiles of 19-mer shRNAs and the corresponding siRNAs. The 19-mer and 29-mer shRNAs and siRNAs designed for seven different target sequences within the coding region of *MAPK14* were tested for gene silencing 24 h after transfection into HeLa cells. Each row of the heat map reports the gene expression signature resulting from transfection of an individual RNA. Two-dimensional clustering of the data groups RNAs (vertical axis dendrogram) and regulated genes (horizontal axis dendrogram) according to signature similarities. Data shown represent genes that display at least a twofold change in expression level ($P < 0.01$ and $\log_{10}$ intensity > 1) relative to mock-transfected cells. Green indicates decreased expression relative to mock transfection and red indicates elevated expression. Black indicates no change and gray indicates data with $P > 0.01$. The red arrow indicates *MAPK14*.

Using microarray gene expression profiling, we obtained a genome-wide view of transcript suppression. A two-dimensional clustering analysis of the signatures produced in HeLa cells 24 h after transfection of 19-mer and 29-mer shRNAs compared with those generated by corresponding siRNAs (**Fig. 4**) shows that each set of three RNAs derived from the same core sequence was accurately clustered. Furthermore, in all but two of seven cases, although the 19-mer shRNAs produced signatures similar to those of the corresponding siRNAs, the signatures of the 29-mer shRNAs were more closely related to those of the corresponding siRNAs. In one of the two cases in which the 19-mer shRNA and the siRNA clustered more closely (*MAPK14-1*), these two RNA species did not appreciably silence the target gene, whereas the 29-mer shRNA did. The agreement between the signatures of 29-mer shRNAs and siRNAs is consistent with precise processing of the shRNA to generate a single siRNA rather than a random sampling of the hairpin stem by Dicer. The overall smaller signature sizes of the 19-mer shRNA and the basis of their divergence from the signature of the corresponding siRNA are presently unclear. However, extensive analysis of off-target effects potentially associated with these shRNAs was not our goal.

Considered together, our results suggest that chemically synthesized 29-mer shRNAs can be substantially more effective triggers of RNAi than can siRNAs. A mechanistic explanation for this finding may lie in the fact that 29-mer shRNAs are substrates for Dicer processing both *in vitro* and *in vivo*. We originally suggested that siRNAs might be passed from Dicer to RISC in a solid-state reaction on the basis of an interaction between Dicer and Argonaute 2 in *D. melanogaster* S2 cell extracts[4]. More recently, results from several laboratories have strongly suggested a model for assembly of the RNAi effector complex in which a multiprotein assembly containing Dicer and accessory proteins interacts with an Argonaute protein and actively loads one strand of the siRNA or miRNA into RISC[18–20]. Such a model implies that Dicer substrates, derived from nuclear processing of pri-miRNAs or cytoplasmic delivery of pre-miRNA mimetics, might be loaded into RISC more effectively than siRNAs. Our data support such a prediction, as it is not the hairpin structure of the synthetic RNA that determines its increased efficacy, but the fact that the shRNA is a Dicer substrate that correlates with enhanced potency, as is reported in an accompanying paper in this issue[29]. In *D. melanogaster*, Dicer is also required for siRNAs to enter RISC, and similar data have been obtained in mammalian cells[18,30]. Thus, it is possible that even siRNAs enter RISC via a Dicer-mediated assembly pathway and that our data simply reflect an increased affinity of Dicer for longer duplex substrates. Alternatively, hairpin RNAs, such as miRNA precursors, might interact with specific cellular proteins that facilitate delivery of these substrates to Dicer, whereas siRNAs might not benefit from such chaperones. Overall, our results suggest an improved method for triggering RNAi in mammalian cells using higher potency RNAi triggers. This remains a critical issue both for cell culture studies and for potential therapeutic use *in vivo*. Mapping the predominant 22-nt sequence that appears in RISC from each of these shRNAs now permits the combination of this more effective triggering method with rules for effective siRNA design.

METHODS

RNA sequence design. Each set of RNAs began with the choice of a single 19-mer sequence. These 19-mers were used directly to create siRNAs. To create shRNAs with 19-mer stems, we appended a four-base loop (either CCAA or UUGG) to the end of the 19-mer sense strand target sequence followed by the 19-mer complementary sequence and a UU overhang. We tested a variety of loop sequences and noted no significant influence of the sequences on the performance of triggers. To create 29-mer stems, we increased the length of the 19-mer target sequence by adding one base upstream and nine bases downstream from the target region and used the same loop sequence and UU overhang. All synthetic RNA molecules used in this study were purchased from Dharmacon.

Dicer processing. RNA hairpins corresponding to luciferase were end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (PNK), and 0.1 pmol of RNA was processed with 2 units of Dicer (Stratagene) at 37 °C for 2 h. Reaction products were Trizol extracted, isopropranol precipitated and separated on an 18% polyacrylamide, 8 M urea denaturing gel. For RNase III digestion, 0.1 pmol was digested with 1 unit of *E. coli* RNase III (NEB) for 30 min at 37 °C and analyzed as described above. Uniformly labeled hairpins were produced using a T7 Megashortscript kit (Ambion) with [α-^{32}P]UTP and then incubated with Dicer as indicated above.

For primer extension analysis, hairpins were processed with Dicer at 37 °C for 2 h; this was followed by heat inactivation of the enzyme. The sequence of the primer is the first 16 nt from the 5′ end of the hairpin: 5′-AGTTGCGCCCGCGAAC-3′. DNA primers were 5′ labeled with PNK and annealed to 0.05 pmol of RNA as follows: 95 °C for 1 min, 10 min at 50 °C and then 1 min on ice. Extensions were carried out at 42 °C for 1 h using MoMLV reverse transcriptase. Products were analyzed by electrophoresis on a 8 M urea/20% polyacrylamide gel.

For analysis of *in vivo* processing, LinxA cells were transfected in 10-cm plates using Mirus TKO (10 μg hairpin RNA) or Mirus LT4 reagent for DNA transfection (12 μg of *AGO1* or *AGO2* DNA)[6]. 293 cells constitutively expressing Ago1 were used for northern blot experiments. Cells were lysed and immunoprecipitated after 48 h using antibody to Myc (9E14). Immunoprecipitates were washed three times in lysis buffer and treated with DNase I for 15 min. Immunoprecipitates were then primer extended as described above.

siRNA and shRNA transfections and mRNA quantification. HeLa cells were transfected in 96-well plates using Oligofectamine (Invitrogen) with the final nanomolar concentrations of each synthetic RNA indicated in the graphs. RNA quantitation was performed by real-time PCR, using appropriate Applied Biosystems TaqMan primer probe sets 24 h after RNA transfection, and the percentage of mRNA remaining was compared with cells treated with transfection reagent alone.

Microarray gene expression profiling. HeLa cells were transfected in six-well plates with 100 nM final concentration of the appropriate RNA using Oligofectamine (according to the manufacturer's instructions). RNA from transfected cells was hybridized competitively with RNA from mock-transfected cells (those treated with transfection reagent in the absence of synthetic RNA). Total RNA was purified by the Qiagen RNeasy kit, and processed as described previously[28] for hybridization to microarrays containing oligonucleotides corresponding to approximately 21,000 human genes. Ratio hybridizations were performed with fluorescent label reversal to eliminate dye bias. Microarrays were purchased from Agilent Technologies. Error models have been described previously[28]. Data were analyzed using Rosetta Resolver software.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

G.J.H. is supported by an Innovator Award from the U.S. Army Breast Cancer Research Program. This work was also supported by a grant from the US National Institutes of Health (G.J.H.). D.S. is supported by a predoctoral fellowship from the US Army Breast Cancer Research Program. We thank the Rosetta Gene Expression Laboratory for microarray RNA processing and hybridizations.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 30 July; accepted 2 November 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
2. Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000).
3. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
4. Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
5. Song, J.J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
6. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
7. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
8. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
9. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
10. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
11. Ketting, R.F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
12. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, 23–34 (2001).
13. Brummelkamp, T.R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
14. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
15. Zeng, Y., Wagner, E.J. & Cullen, B.R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
16. Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
17. Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
18. Lee, Y.S. *et al.* Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
19. Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. & Sontheimer, E.J. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**, 83–94 (2004).
20. Tomari, Y. *et al.* RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**, 831–841 (2004).
21. Zhang, H., Kolb, F.A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
22. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
23. Ma, J.B., Ye, K. & Patel, D.J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
24. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**, 465–469 (2003).
25. Song, J.J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* **10**, 1026–1032 (2003).
26. Yan, K.S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
27. Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
28. Jackson, A.L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
29. Rossi, J.J. *et al.* Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.* **23**, in the press (2005).
30. Doi, N. *et al.* Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr. Biol.* **13**, 41–46 (2003).

Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis

Michele A Cleary[1], Kristopher Kilian[1], Yanqun Wang[1], Jeff Bradshaw[1], Guy Cavet[1], Wei Ge[1], Amit Kulkarni[1], Patrick J Paddison[2], Kenneth Chang[2], Nihar Sheth[2], Eric Leproust[3], Ernest M Coffey[1], Julja Burchard[1], W Richard McCombie[2], Peter Linsley[1] & Gregory J Hannon[2]

Generation of complex libraries of defined nucleic acid sequences can greatly aid the functional analysis of protein and gene function. Previously, such studies relied either on individually synthesized oligonucleotides or on cellular nucleic acids as the starting material. As each method has disadvantages, we have developed a rapid and cost-effective alternative for construction of small-fragment DNA libraries of defined sequences. This approach uses *in situ* microarray DNA synthesis for generation of complex oligonucleotide populations. These populations can be recovered and either used directly or immortalized by cloning. From a single microarray, a library containing thousands of unique sequences can be generated. As an example of the potential applications of this technology, we have tested the approach for the production of plasmids encoding short hairpin RNAs (shRNAs) targeting numerous human and mouse genes. We achieved high-fidelity clone retrieval with a uniform representation of intended library sequences.

Nucleic acid libraries provide some of the most versatile tools for functional analysis of genomes, individual proteins or complexes[1–14]. These libraries can be constructed using either biologically derived or chemically synthesized nucleic acids as substrates. Libraries generated from natural sources generally do not cover all expressed sequences in the genome, largely owing to tissue-specific mRNA expression and variations in mRNA abundance, limiting the complexity and uniform representation of the cDNA source material. Chemically synthesized oligonucleotides have also been used to construct libraries for biological analysis[14]. Although these allow defined and uniform representation, the cost of source material for library construction is quite high.

To reduce the cost inherent in the use of conventional methods for the generation of complex libraries of defined nucleic acids, we have developed an approach that uses printed microarrays as a source material for complex oligonucleotide populations[15–18] (**Fig. 1**). Although such an approach can be applied in many different ways, we have tested the methodology for one specific application, namely for the construction of libraries of shRNA expression constructs.

RESULTS

Ink-jet synthesis of oligonucleotides

Ink-jet technology has been optimized for hybridization microarrays using oligonucleotides of 60 bases or less on slides that contain ~25,000 individual spots[15–18]. But no tests have suggested whether this method produced DNA of sufficient quality or quantity for use as source material for library construction. To address this question, we designed and printed arrays containing 110 unique 59-nucleotide (nt) DNA sequences, each containing identical flanking PCR primer binding sites. Initially, each oligonucleotide was synthesized redundantly in ~220 different locations on an array containing 24,200 probes to give an overall complexity of 1,000 different sequences. Oligonucleotide populations were recovered from the microarray surface using one of two approaches. The first, simpler approach involved treatment of standard arrays with ammonium hydroxide[19] (**Fig. 1a**). The second approach required derivatizing slides with a photocleavable linker before synthesis, and the oligonucleotides were ultimately recovered after a brief treatment with UV light. After harvesting the oligonucleotides, we amplified the pooled material by PCR and cloned the products. Of the clones obtained from ammonium hydroxide cleaved material, five of five readable sequences were of the correct length, exactly matched one of the sequences in the array pattern design, and were unique. Of the clones obtained from photocleaved material, four of five readable sequences had the correct length and each perfectly matched a unique sequence in the array pattern design. These results suggested that the use of this highly parallel synthesis approach was feasible for producing clones with ~60-base-pair (bp) inserts.

Accurate synthesis of long oligonucleotides

The ability to produce complex libraries comprised of defined 60-nt fragments is sufficient for many purposes, and such arrays can be purchased as a standard product from Agilent Technologies, the

[1]Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., 401 Terry Ave. North, Seattle, Washington 98109, USA. [2]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [3]Agilent Technologies, 3500 Deer Creek Road, Palo Alto, California 94304, USA. Correspondence should be addressed to M.A.C. (Michele_Cleary@merck.com) and G.J.H. (hannon@cshl.edu).

PUBLISHED ONLINE 18 NOVEMBER 2004; DOI:10.1038/NMETH724

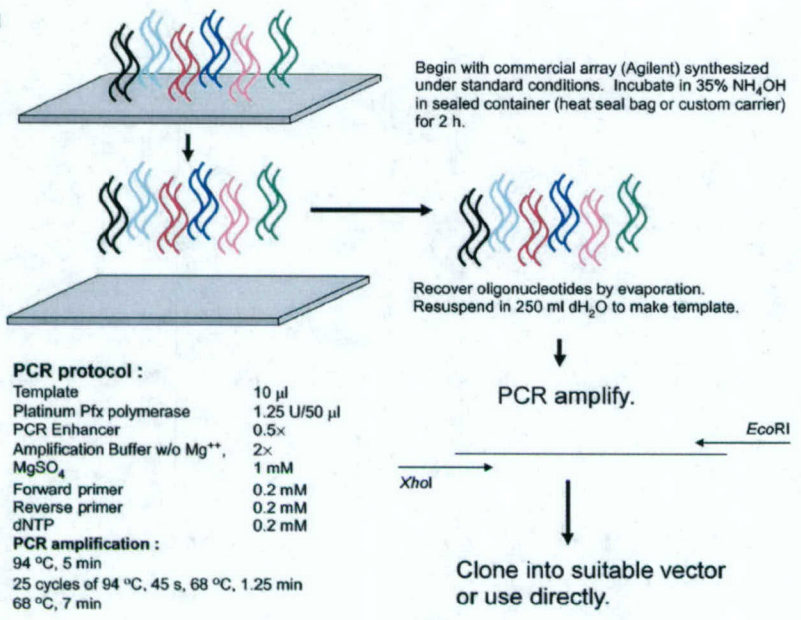


Figure 1 | Cloning strategy using *in situ* oligonucleotide synthesis. To create a pool of sequences for library cloning, oligonucleotides were printed on a microarray substrate, cleaved by treatment with a strong base or ultraviolet light and amplified by PCR. The amplified products were treated with restriction enzymes or used directly for ligation as a pool into a vector of choice.

commercial source of arrays used in this report. However, some specialized applications may require longer oligonucleotides. For our purposes, the design of optimized shRNA libraries requires synthesis of oligonucleotides ~100 bases in length. Although these are not a standard product, the use of 100-mers provided a very stringent test of the array methodology for library production. We designed arrays containing 96-nt sequences deposited either once per array or at variable representation, ranging from 1 to 1,024 times. With PCR products derived from ammonium hydroxide–cleaved material, we found an average of ~63% of clones (total of 30 in three separate cloning trials) with the correct sequence and length (**Fig. 1b**, primer and template structure). From arrays printed at variable oligonucleotide representation, we recovered an overwhelming majority of accurate clones corresponding to the sequence spotted 1,024 times. We could not clone from photo-cleaved 96-nt material.

The use of RNA interference (RNAi) has opened the door for loss-of-function genetic approaches in numerous organisms, including mammals[20]. One method to achieve RNAi is the expression of shRNAs from DNA vectors[21–27]. We therefore set out to use *in situ*–synthesized sequences to build shRNA expression libraries targeting nearly every identified and predicted gene in the genomes of several species, including human, mouse and rat. Similar libraries have previously been constructed using conventional oligonucleotides or natural nucleic acids as starting material[28–33]. To maximize recovery of accurate clones from our highly structured templates, we used thermostable polymerases that have proofreading capability and are able to effect strand displacement. We also added PCR enhancing agents such as DMSO or betaine. Through a combination of these strategies (**Fig. 1** and Methods), we were able to achieve success rates consistently ranging from ~25% to >60% for cloning of perfect shRNAs.

Construction of shRNA libraries

In our effort to create large-scale human and mouse shRNA libraries, we designed oligonucleotides corresponding to more than 32,000 known and predicted genes each in human and mouse. These yield 195,077 oligonucleotides homologous to murine genes and 187,905 oligonucleotides homologous to human genes. Each oligonucleotide was synthesized once on each array, necessitating the use of a minimum of 21 arrays to completely cover genes in both organisms with up to six shRNAs each. Iterative cycles of sequencing and synthesis were used to maximize the efficiency of obtaining correct clones. Recovery of unique, perfect shRNA vectors from the population can be hampered by two types of errors. The first is inaccuracies in the synthesis, amplification or sequencing leading to inserts that are or appear inaccurate. The second is biases in the synthesis, amplification and cloning procedures leading to imperfect representation of the desired oligonucleotide population in the cloned pools. We have examined each problem separately. **Table 1** shows data relevant to the first type of error, comparing the accuracy of array synthesis and chemical synthesis. Thus far, we have sample-sequenced clones from 23 separate arrays covering a total of 447,410 printed sequences with 216,945 informative sequencing reads. An informative read is defined as a sequencing run that gives high-quality sequence (PHRED score >20 over the length of the insert). The rates of successfully obtaining perfect clones varied from 21% to 58%, depending on the synthesis run. The ~220,000 reads yielded 76,960 perfect clones overall, of which 53,478 represented unique sequences. We noted no bias for correct versus incorrect clones based on the oligonucleotide position on the arrays. For comparison, we obtained 7,360 oligonucleotides in six independent batches that were produced using conventional synthesis methods by a commercial manufacturer. From 18,554 informative sequencing reads, 3,526 perfect clones were obtained, with success rates from individual pools ranging from 9.9% to 26.5%. These pools represent the upper range of success rates with conventional oligonucleotides obtained from a number of different suppliers. As it is difficult to directly compare the quality of array synthesized material to conventionally synthesized material given differences in pool complexities, it seems reasonable to conclude that the array synthesized material, treated and cloned in the manner described herein, is of a quality that is at least equivalent to that of conventionally synthesized material.

Table 1 | Characterization of cloned shRNAs

Source	Complexity	Reads	Correct	1 Mm	2 Mm	>2 Mm	Success rate
In situ–synthesized oligonucleotides							
Chip 1	19,253	16,103	3,671	3,722	2,050	6,660	0.227
Chip 2	19,230	12,531	3,749	3,962	1,782	3,038	0.299
Chip 3	19,244	6,877	2,533	1,859	728	1,757	0.368
Chip 4	19,245	12,749	3,312	3,321	1,878	4,238	0.259
Chip 5	19,251	10,852	5,232	2,004	527	3,089	0.482
Chip 6	19,234	13,079	3,847	3,703	1,594	3,935	0.294
Chip 7	19,216	12,616	3,260	3,861	1,934	3,561	0.258
Chip 8	19,226	12,584	4,236	3,805	1,630	2,913	0.336
Chip 9	19,228	12,542	2,635	4,052	2,312	3,543	0.21
Chip 10	22,089	7,550	3,880	1,436	412	1,822	0.513
Chip 11	19,236	7,402	4,342	1,533	386	1,141	0.586
Chip 12	22,077	7,351	3,596	1,447	449	1,859	0.489
Chip 13	21,524	7,502	3,775	1,542	415	1,770	0.503
Chip 14	22,000	5,826	1,439	1,791	847	1,749	0.246
Chip 15	17,621	11,911	6,579	2,598	582	2,152	0.552
Chip 16	20,545	6,030	2,626	1,611	416	1,377	0.435
Chip 17	20,550	5,534	1,645	1,769	633	1,487	0.297
Chip 18	20,546	14,229	3,279	4,264	2,268	4,418	0.23
Chip 19	17,620	5,425	1,783	1,785	573	1,284	0.328
Chip 20	17,620	5,756	2,007	1,828	585	1,336	0.348
Chip 21	17,616	5,503	2,013	1,758	517	1,215	0.365
Chip 22	17,621	11,600	5,087	2,678	840	2,995	0.438
Chip 23	17,618	5,393	2,434	1,399	356	1,204	0.451
Conventionally synthesized oligonucleotides							
Pool A	576	1,827	396	240	120	1,071	0.216
Pool B	2,000	6,751	1,792	1,189	645	3,125	0.265
Pool C	2,000	2,726	180	302	173	2,071	0.066
Pool D	1,440	3,851	695	529	305	2,322	0.18
Pool E	768	1,510	150	183	159	1,018	0.099
Pool F	576	1,889	313	426	224	926	0.165

Oligonucleotides from the indicated sources were amplified by PCR and cloned into pSM2 (G.J.H., J. Silva, P.J.P., M.A.C., S. Elledge, D. Siolas *et al.*, data not shown) for sequencing. The complexity of each population is indicated. Informative reads (Reads) were scored for perfect clones (Correct) or for clones with a single mismatch (1 Mm), two mismatches (2 Mm) or more than two mismatches (>2 Mm). An informative read is defined as a trace giving a sequence of sufficient quality for analysis (PHRED score >20 over the length of the insert). The success rate is calculated as the fraction of perfect clones from the total informative reads.

Table 2 presents data that tracks the second type of error, measuring the frequency with which we recovered individual oligonucleotide sequences as clones. To examine the data in the most consistent fashion, we examined the performance of each pool when the sampling by sequencing had reached 0.5×. We scored all identifiable clones, defined as those with a sequence with fewer than three mismatches to the target. Overall, pools made by both synthesis methods behaved similarly. Both sources of material yielded clone populations that matched slightly fewer oligonucleotides than was expected from a Poisson distribution, indicating that there were inherent biases in either the synthesis or the amplification of each oligonucleotide population. With conventionally synthesized material the rate at which cloned oligonucleotides were recovered in a nonredundant fashion varied from 34% to 68%, whereas with array-synthesized material this varied from 51% to 70%. At 0.5× sampling, ~78% of reads were expected to represent unique oligonucleotides.

To examine a single population as an example, consider chip 15 (**Tables 1** and **2**). Of 11,911 informative reads, 6,579 perfectly matched printed oligonucleotide sequences, giving an accuracy rate of 55.2%. To measure sampling error, we considered only the first 8,810 reads that unambiguously matched oligonucleotides printed on the array so that our sampling rate was normalized with other populations at ~0.5×. Within those 8,810 reads, we expected that 6,933 printed oligonucleotide sequences would be represented (78% of 8,810). Instead, we found 4,780 printed sequences.

An examination of the melting temperature (T_m) profile of the recovered, perfect shRNAs showed that it largely reflected the T_m profile of the total library oligonucleotide population, although there was a shift toward lower T_m for perfect clones (**Fig. 2a**). Similar results were obtained for conventional oligonucleotides (**Fig. 2b**). These results suggested that the PCR and cloning procedures used had a small preference for amplification of hairpins with lower thermal stability. The difference in T_m between the perfect and expected clones represents a shift corresponding to approximately two additional G-C base pairs. Furthermore, an examination of the error profile of the sequences suggested that there exists a bias for errors within the stem regions (**Fig. 2c,d**). This same bias was seen irrespective of the source of the oligonucleotides, with conventional and ink-jet samples giving similar results.

Table 2 | Sampling of shRNA populations from Chip and conventional oligonucleotides

Source	Complexity	Reads	Sampling rate	Expected unique	Actual unique	Unique rate
In situ–synthesized oligonucleotides						
Chip 1	19,253	9,626	0.499	7,575	6,821	0.708
Chip 2	19,230	9,615	0.5	7,566	5,617	0.584
Chip 4	19,245	9,622	0.499	7,572	6,439	0.669
Chip 5	19,251	9,625	0.499	7,574	5,756	0.598
Chip 6	19,234	9,617	0.5	7,567	5,694	0.592
Chip 7	19,216	9,608	0.5	7,560	6,116	0.636
Chip 8	19,226	9,613	0.5	7,564	5,521	0.574
Chip 9	19,228	9,614	0.5	7,565	5,745	0.597
Chip 15	17,621	8,810	0.499	6,933	4,780	0.542
Chip 18	20,546	10,273	0.5	8,084	7,330	0.713
Chip 22	17,621	8,810	0.499	6,933	4,529	0.514
Conventionally synthesized oligonucleotides						
Pool A	576	288	0.5	226	197	0.684
Pool B	2,000	1,000	0.5	786	575	0.575
Pool C	2,000	1,000	0.5	786	489	0.489
Pool D	1,440	720	0.5	566	480	0.666
Pool E	768	384	0.5	302	235	0.611
Pool F	576	288	0.5	226	99	0.343

For all populations sampled at more than 0.5× coverage, we determined the redundancy rate in the informative reads. So that each population could be compared, we restricted our analysis to sequencing reads that comprised only 0.5× coverage. The actual and expected unique matches to printed oligonucleotides for each pool, according to a Poisson distribution, are given along with the rate at which unique matches are obtained.

The peaks of errors that are observed within the loop region do not correspond to any regions of known structure. All represent adenine residues, however, potentially indicating some bias in the chemical synthesis procedure.

Array-based assessments of synthesis bias

To assess the representation of the printed sequences in the amplified oligonucleotide pools, we used standard microarray hybridization. We printed and cleaved a set of 18,723 unique 97-base oligonucleotides encoding shRNAs each spotted once on the array. We also designed four subset arrays, each containing 5,152 of the 18,723 sequences, with each subset overlapping the subsequent subset by ~600 sequences. We used a T7 promoter–adapted PCR primer to amplify double-stranded templates for *in vitro* transcription (IVT), transcribed these templates in the presence of amino allyl UTP and coupled the resulting IVT

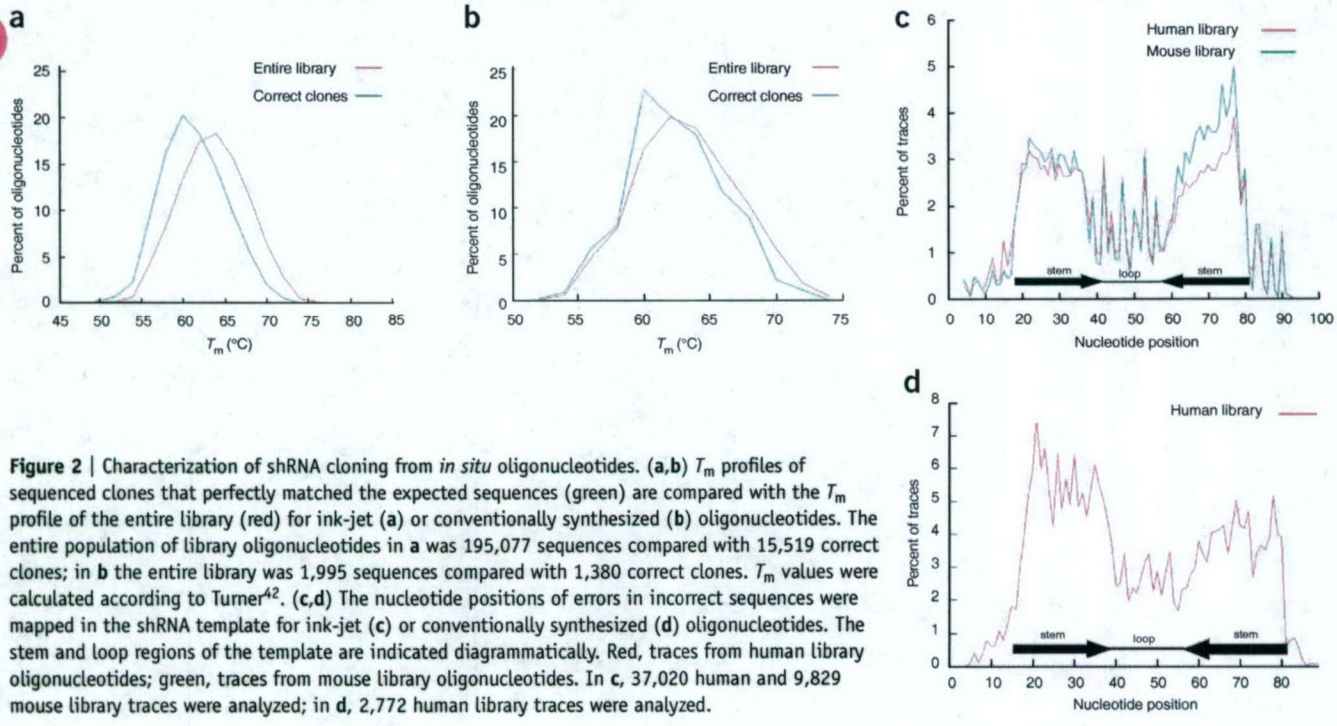


Figure 2 | Characterization of shRNA cloning from *in situ* oligonucleotides. (**a**,**b**) T_m profiles of sequenced clones that perfectly matched the expected sequences (green) are compared with the T_m profile of the entire library (red) for ink-jet (**a**) or conventionally synthesized (**b**) oligonucleotides. The entire population of library oligonucleotides in **a** was 195,077 sequences compared with 15,519 correct clones; in **b** the entire library was 1,995 sequences compared with 1,380 correct clones. T_m values were calculated according to Turner[42]. (**c**,**d**) The nucleotide positions of errors in incorrect sequences were mapped in the shRNA template for ink-jet (**c**) or conventionally synthesized (**d**) oligonucleotides. The stem and loop regions of the template are indicated diagrammatically. Red, traces from human library oligonucleotides; green, traces from mouse library oligonucleotides. In **c**, 37,020 human and 9,829 mouse library traces were analyzed; in **d**, 2,772 human library traces were analyzed.

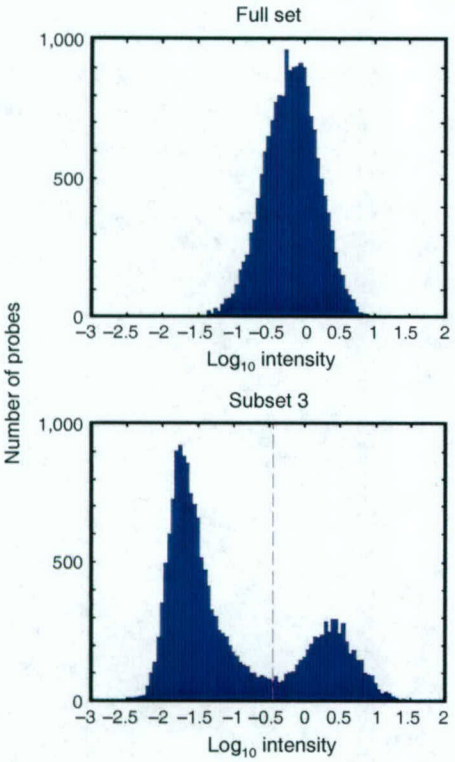


Figure 3 | Histograms of the average intensity of the 18,723 probes when hybridized to IVT products derived from the pool of the full-set of sequences (top) and one representative subset of 5,152 sequences (bottom). Subset arrays 1, 2 and 4 showed similar bimodal distributions.

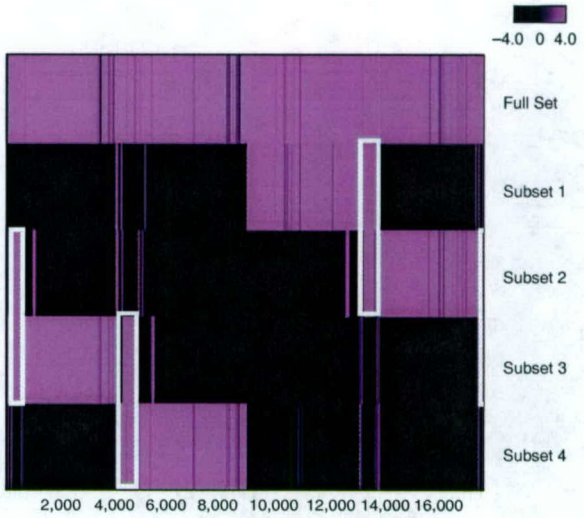


Figure 4 | The subset sequences gave unique signatures of bright-intensity probes and showed the expected overlap. The heat map shows the results of two-dimensional clustering of logarithmic intensities of 17,552 good probes, representing >98% of the 17,898 valid probes (excluding 825 total GGGTTGGCTC-containing sequences) on the full set and subset cloning array samples. Pink, bright-intensity probes; black, dim-intensity probes; white boxes, probes with expected overlap among the subset arrays. Note that the probe intensity from each array is normalized by its computed threshold for representation so that a sequence is considered represented when its logarithmic intensity is >0.

products to Cy3 and Cy5 dyes. After coupling, we hybridized dye-labeled material to a 'diagnostic' microarray containing 60-mer probes of all 18,723 sequences, along with controls. To minimize cross-hybridization, we eliminated the common primer binding sites from the oligonucleotides on the diagnostic array. In these shRNAs, up to three G-C base pairs in the stems were converted to encode G-U base pairs in the expressed shRNAs[28]. This approach alleviates secondary structure at the DNA level and increases stability during amplification, cloning and propagation in bacteria. Newer shRNA designs, such as those used for the sequence analysis of shRNA populations described above, do not incorporate this strategy, but the inclusion of G-U mismatches in the stem region should have no impact on the relative degree to which cleaved populations represent the total pool of synthesized material.

We observed a single-mode distribution of hybridizing probes (high and low intensity) on the diagnostic microarray for the full-set pool and, as expected, bimodal distributions for the subset pools (**Fig. 3**). After subtraction of background hybridization using negative controls on the microarray, the distributions were segmented to estimate the probes with intensity above background as follows. For hybridization to the subset pools, we used the data from the subset detection arrays to calculate false positive and false negative rates. A false positive for a subset array is a sequence determined to be represented in the hybridization but not included in the 5,152 sequences actually printed on the array from which the pool was derived. A false negative is a sequence that was not represented in the hybridization, despite being an intended sequence of the set. For each subset array, the threshold for representation was set such that the sum of the false positive rate and the false negative rate was minimized. The computed threshold essentially segments the bimodal probe intensity distribution into two groups, represented sequences and background (**Fig. 3**). The same approach can be extended to the full-set array to estimate the number of sequences deemed represented, in which case the representation threshold segments the full-set probes (represented) from the negative control probes (background).

By this approach, labeled IVT products from the full-set of sequences hybridized to 18,686 (99.8%) of the 18,723 unique sequence probes. The collective data for the four subset oligonucleotide pools revealed 390 sequences that overlapped in all four hybridization experiments. This overlap was not intended in the array design. On further inspection, it became apparent that members of this set of sequences shared a highly conserved internal core of approximately ten consecutive bases (GGGTTGGCTC) that included the conserved shRNA loop structure (**Supplementary Fig. 1** online). These fortuitous stretches of sequence conservation likely explain the cross-hybridization observed. Of the probes on the microarray, 825 sequences contain the sequence GGGTTGGCTC from positions 27–36.

As a visual illustration of the coverage afforded by our library pools, we eliminated the 825 probes with the common core sequence GGGTTGGCTC and studied only the 17,898 remaining valid probes. Using the segmentation method described earlier, we obtained 17,552 probes with hybridization intensity substantially above background in at least one subset detection array (representing more than 98% of the 17,898 valid probes) and carried out a two-dimensional intensity clustering analysis of these probes. Each

cleaved subset array gave a unique signature (**Fig. 4**). As expected, we observed small clusters of bright probes for each array that were also bright for intended overlapping arrays (white boxes). With this approach, we obtained an average false positive rate of 6.15% and an average false negative rate of 1.99%. The higher, but still quite low, false positive rate likely reflects a much smaller set of redundant sequences that remains after removal of the 825 GGGTTGGCTC-containing sequences (data not shown). Thus, the true false positive rate probably approaches that of the false negative rate. Considered together with the sample sequencing, these data suggest that pools of oligonucleotides cleaved from microarrays are well represented.

DISCUSSION

Cost-effective approaches for cloning complex libraries of predefined nucleic acid sequences are very limited. Typically, if there is no natural source of the nucleic acid, oligonucleotides must be synthesized individually for engineering into the larger library. This traditional approach is disadvantageous in several respects. First, it is costly, which limits the number of sequences that can be included in the library. Second, the approach is labor intensive, as each individual oligonucleotide must be manipulated for engineering into the library. Even in cases where natural sources of nucleic acid are available, cloning and manipulation of these might not produce ideally structured populations. Our data show that microarray-based library cloning provides a rapid, cost-effective and flexible approach for the generation of complex, uniformly distributed libraries of defined oligonucleotides.

Ink-jet microarray synthesis has been optimized for production of oligonucleotides of 60 bases or fewer, and such standard arrays will be suitable for many purposes. We have shown that we can use ink-jet synthesis to produce very high-fidelity cloned populations with oligonucleotides of up to 96 bases. Although arrays carrying oligonucleotides of this length are not standard reagents, we used these materials to provide a very stringent test for the performance of array-synthesized oligonucleotides. We noted high fidelity and only modest biases in the amplification of complex populations of highly structured templates. Overall, considering only the accuracy of cloned populations, we consistently recovered 45–55% of clones with perfect sequences. Also considering biases in amplified populations, 25–30% of all clones represent unique and perfect shRNAs. Both the rates themselves and the importance of each metric will vary with individual applications of the approach. For our specific purpose, success rates in generating viable shRNA clones using ink-jet-synthesized oligonucleotides are sufficient to allow this method to be used for the large-scale construction of both mixed and sequence verified libraries, as it does not substantially differ from success rates observed in our previous efforts at library construction using conventionally synthesized material[28].

The creation of complex libraries by ink-jet DNA synthesis can be applied to address numerous biological problems. For example, this method would be ideal for generating libraries for antibody diversity studies, phage display, combinatorial peptide sequence generation, DNA binding site selection, promoter region analysis and restriction enzyme site analysis. In each case, the necessary oligonucleotide length, the requirement for sequence verification and the arraying clones will vary. Accordingly, the cost savings afforded by this technology will also vary. Overall, our data suggest a parity between the quality of ink-jet synthesized material and material obtained by mixing populations of conventional oligonucleotides. Given the accuracy and flexibility of ink-jet oligonucleotide synthesis, it is likely that the approach described here will become an important method for constructing diverse library-based tools for functional genomic studies.

METHODS

Oligonucleotide design and microarray synthesis. For this cloning method, any microarray technology capable of *in situ* synthesis of oligonucleotides of the desired length for the application may be appropriate. For our studies, however, we primarily used and validated oligonucleotide microarrays printed at Agilent Technologies using ink-jet technology as described previously[15] with essentially no modifications to standard manufacturer's protocols. Detailed methods for generating ink-jet microarrays can be found in U.S. Patents numbered 6,419,883 and 6,028,189.

Sequences to be included in a library were designed such that each was flanked by 5' and 3' common 14- to 18-base PCR primer recognition sites (**Fig. 1b**). Before the oligonucleotides were harvested, quality control testing was performed using a functional hybridization of representative arrays that were produced on the same manufactured glass substrates.

Oligonucleotide cleavage with a photocleavable spacer. Photocleavable spacer phosphoramidite (Glen Research) monomers were synthesized on a silanized 3 inch × 3 inch × 0.004 inch glass wafer with hydroxyl functionality. Silanization of glass surfaces for oligonucleotide applications has been described[30,32] and silanes with various functionalities are commercially available (Gelest). For these studies a 50:1 mixture of decyl trichlorosilane and 11-trichlorosilyl-1-undecene was used. All reaction steps and reagent preparations were performed under nitrogen in a PLAS-LABS 830-ABC glove box (PLAS-LABS). One microliter of anhydrous acetonitrile (Fisher Scientific) was added by syringe injection to 100 μmol of freeze-dried photocleavable spacer phosphoramidite to yield a 0.1 M solution. Next, 62 ml of anhydrous acetonitrile was added to 2 g of freeze-dried 5-ethylthiol-^{1}H-tetrazole (Glen Research) to yield a 0.25 M solution for phosphoramidite activation. The solutions were vortexed briefly and allowed to equilibrate at room temperature for 30 min. One milliliter of tetrazole solution was transferred by syringe to the photocleavable spacer solution, and the mixture was vortexed for 10 s. Two silanized wafers were placed 'reactive side' up and 2 ml of the active photocleavable spacer–tetrazole solution was added to the surface of the first wafer. The second wafer was placed sandwich-like on top of the first, allowing the fluid to distribute uniformly between the surfaces. The wafers were incubated at room temperature for 2 min, separated, placed in a Teflon rack and immersed in a bath of acetonitrile. The rack was agitated in the bath for 2 min to ensure complete rinsing of excess photocleavable spacer and dried by centrifugation. Formation of the stable pentavalent phosphodiester and removal of the dimethoxytrityl protecting group were carried out according to standard oligonucleotide synthesis procedures[15,17]. Synthesis of oligonucleotides on photocleavable spacer–functionalized substrate was performed as described above.

For arrays synthesized with a photocleavable spacer, the oligonucleotides were cleaved in 1 ml of 25 mM Tris-buffer solution (pH 7.4) by placing the array in almost direct contact with a UV

irradiation source (UVM–57, UVP, Inc.; 302 nm wavelength) for 20 min. The solution was transferred to a 1.5-ml microcentrifuge tube and speed vacuumed at 45 °C overnight.

Oligonucleotide cleavage using ammonium hydroxide. To cleave oligonucleotides synthesized without a photocleavable spacer, the microarrays were treated for 2 h with 2–3 ml of 35% NH_4OH solution (Fisher Scientific) at room temperature. The solution was transferred to 1.5-ml microcentrifuge tubes and speed vacuum dried at 45 °C overnight.

PCR amplification of cleaved oligonucleotides. Dried material containing oligonucleotides cleaved from each microarray was resuspended in 250 µl of RNase- and DNase-free water. For the PCR template, a range of volumes (0.1–5.0 µl) was tested to determine the amount that gave the best yield with the lowest incidence of nonspecific product. We carried out PCR amplification of the initial 59- and 96-nt test sequences in 50 µl reactions containing 1x PCR buffer without Mg (Invitrogen), 9% sucrose, 1.5 mM $MgCl_2$, 1 ng/µl forward and reverse primers, 125 µM dNTPs and 0.05 U/µl *Taq* polymerase. Thermal cycler conditions depended on the length of the oligonucleotides and the melting temperatures of the forward and reverse primers. In general, 30 cycles of 94 °C denaturing for 30 sec, annealing at the appropriate temperature for 30 s, and extension at 72 °C for 90 s worked well. If the PCR products were to be cloned using a TA cloning system such as the Topo TA cloning system (Invitrogen), we used *Taq* polymerase and followed the 30-cycle PCR with a 10 min extension at 72 °C. For the cloning of shRNA libraries, Vent polymerase (New England Biolabs) or Pfx polymerase (Invitrogen) in the presence of DMSO and/or betaine was used to reduce the incidence of nucleotide misincorporation during the PCR. We optimized conditions independently for each primer set used. For the cloning of mouse and human shRNA libraries discussed here, we used Platinum Pfx (Invitrogen) with a 2× final concentration of the manufacturer's provided amplification buffer, a 0.5× final concentration of the provided PCR enhancer, 10 µl of template (1/50 of the oligonucleotides cleaved from each array), 1 mM $MgSO_4$, 0.2 µM each of the forward and reverse primers, and 0.2 mM final concentration of all four dNTPs. Thermal cycler conditions were 94 °C for 5 min followed by 25 cycles of 94 °C for 45 s, 68 °C for 1 min 15 s, and finally extension at 68 °C for 7 min. In some cases, PCR products were cleaned up by gel purification using the QIAquick Gel Extraction protocol (QIAgen). In other cases, the PCR products were cleaned up using the QIAquick PCR purification protocol (QIAgen).

Reverse transcription–*in vitro* transcription (RT-IVT) and microarray hybridization. To prepare templates for T7 IVT, we pooled PCR material from two individual reactions. Unincorporated nucleotides and polymerase were removed from the pooled PCR products by QIAquick PCR purification (QIAgen) and eluted in 50 µl of RNase and DNase-free water. Eluates were speed-vacuum dried to concentrate two-fold and 7.25 µl was used as template in a T7 RNA polymerization reaction, using a modified MEGAshortscript protocol (Ambion). In lieu of 2 µl of 75 mM UTP, we used 2.25 µl of 50 mM amino allyl UTP (aa-UTP; Ambion) plus 0.5 µl of the 75 mM UTP provided with the kit. The reactions were carried out at 37 °C overnight. Then, 1 µl of DNase was added for 15 min at room temperature. Next, the samples were phenol/chloroform/isoamyl alcohol extracted and ethanol precipitated. The final product was resuspended in 40 µl of water.

Amino allyl UTP–incorporated cRNA was divided into aliquots in two 96-well plates (5 µg per reaction well). One plate for Cy3 NHS-ester coupling and one for Cy5 NHS-ester coupling were prepared (dyes were obtained from Amersham Biosciences). Samples were reacted with the dyes, mixed for performance of two-color ratio experiments and subsequently purified using Micro Bio-Spin columns P-30 Tris (Bio-Rad Laboratories). Purified dye-labeled samples were then hybridized to the detection microarray for 24 h, washed, scanned on an Agilent Scanner and analyzed. Rosetta standard coupling and hybridization processes were employed as previously described[15].

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

G.J.H. is supported by an Innovator Award from the US Army Breast Cancer Research Program. This work was also supported by a grant from the US National Institutes of Health (G.J.H.). We thank H. Dai for suggestions regarding microarray analysis of the library population, T. Fare for helpful comments on the manuscript, and the Rosetta Gene Expression Laboratory for microarray RNA processing and hybridizations.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 26 July; accepted 20 October 2004
Published online at http://www.nature.com/naturemethods/

1. Hudson, J.D. *et al.* A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J. Exp. Med.* **190**, 1375–1382 (1999).
2. Brummelkamp, T.R. *et al.* TBX-3, the gene mutated in Ulnar-Mammary Syndrome, is a negative regulator of p19ARF and inhibits senescence. *J. Biol. Chem.* **277**, 6567–6572 (2002).
3. Maestro, R. *et al.* Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* **13**, 2207–2217 (1999).
4. Raveh, T., Berissi, H., Eisenstein, M., Spivak, T. & Kimchi, A. A functional genetic screen identifies regions at the C-terminal tail and death-domain of death-associated protein kinase that are critical for its proapoptotic activity. *Proc. Natl. Acad. Sci. USA* **97**, 1572–1577 (2000).
5. Fletcher, B.S., Dragstedt, C., Notterpek, L. & Nolan, G.P. Functional cloning of SPIN-2, a nuclear anti-apoptotic protein with roles in cell cycle progression. *Leukemia* **16**, 1507–1518 (2002).
6. Rayner, J.R. & Gonda, T.J. A simple and efficient procedure for generating stable expression libraries by cDNA cloning in a retroviral vector. *Mol. Cell. Biol.* **14**, 880–887 (1994).
7. Kitamura, T. *et al.* Efficient screening of retroviral cDNA expression libraries. *Proc. Natl. Acad. Sci. USA* **92**, 9146–9150 (1995).
8. Kojima, T. & Kitamura, T. A signal sequence trap based on a constitutively active cytokine receptor. *Nat. Biotechnol.* **17**, 487–490 (1999).
9. Golovkina, T.V. *et al.* A novel membrane protein is a mouse mammary tumor virus receptor. *J. Virol.* **72**, 3066–3071 (1998).
10. Battini, J.L., Rasko, J.E. & Miller, A.D. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: possible role in G protein-coupled signal transduction. *Proc. Natl. Acad. Sci. USA* **96**, 1385–1390 (1999).
11. Gallagher, W.M., Cairney, M., Schott, B., Roninson, I.B. & Brown, R. Identification of p53 genetic suppressor elements which confer resistance to cisplatin. *Oncogene* **14**, 185–193 (1997).
12. Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nat. Genet.* **14**, 415–420 (1996).
13. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. & Toole, J.J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).
14. Tuerk, C., MacDougal, S. & Gold, L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **89**, 6988–6992 (1992).

15. Hughes, T.R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).
16. Halliwell, C. & Cass, A.E. A factorial analysis of silanization conditions for the immobilization of oligonucleotides on glass surfaces. *Anal. Chem.* **73**, 2476–2483 (2001).
17. Brown, D.M. A brief history of oligonucleotide synthesis. *Methods Mol. Biol.* **20**, 1–17 (1993).
18. Bourdeiu, L., Silberzan, P. & Chatenay, D. Langmuir-Blodgett films: From micron to angstrom. *Physical Review Letters* **7**, 2029–2032 (1991).
19. LeProust, E., Zhang, H., Yu, P., Zhou, X. & Gao, X. Characterization of oligodeoxyribonucleotide synthesis on glass plates. *Nucleic Acids Res.* **29**, 2171–2180 (2001).
20. Hannon, G.J. & Rossi, J.J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
21. Brummelkamp, T.R., Bernards, R. & Agami, R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **296**, 550–553 (2002).
22. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
23. Paul, C.P., Good, P.D., Winer, I. & Engelke, D.R. Effective expression of small interfering RNA in human cells. *Nat. Biotechnol.* **20**, 505–508 (2002).
24. Lee, N.S. *et al.* Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat. Biotechnol.* **20**, 500–505 (2002).
25. Sui, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA* **99**, 5515–5520 (2002).
26. Kawasaki, H. & Taira, K. Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res.* **31**, 700–707 (2003).
27. Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* **20**, 497–500 (2002).
28. Paddison, P.J. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
29. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
30. Sen, G., Wehrman, T.S., Myers, J.W. & Blau, H.M. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat. Genet.* **36**, 183–189 (2004).
31. Shirane, D. *et al.* Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36**, 190–196 (2004).
32. Luo, B., Heard, A.D. & Lodish, H.F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc. Natl. Acad. Sci. USA* **101**, 5494–5499 (2004).
33. Hsieh, A.C. *et al.* A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32**, 893–901 (2004).

PUBLISHED IN ASSOCIATION WITH
COLD SPRING HARBOR LABORATORY

Cloning of short hairpin RNAs for gene knockdown in mammalian cells

Patrick J Paddison[1], Michele Cleary[2], Jose Maria Silva[1], Kenneth Chang[1], Nihar Sheth[1], Ravi Sachidanandam[1] & Gregory J Hannon[1]

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Rosetta Inpharmatics, 401 Terry Avenue N., Seattle, Washington 98109, USA.

RNA interference (RNAi) has become a methodology of choice for knocking down gene expression in a variety of biological systems[1,2]. The demonstration, in mammalian *in vitro* systems, of gene silencing using double-stranded RNA (dsRNA) products <30 base pairs (bp) in length has placed RNAi at the forefront of gene manipulation techniques in somatic cells[3–6]. Two types of dsRNA triggers are now commonly used to evoke RNAi in mammalian cells: (i) chemically or *in vitro*–synthesized small interfering RNAs (siRNAs)[3,4] and (ii) short hairpin RNAs (shRNAs) expressed from RNA polymerase III promoters[2]. We and others have chosen to explore the shRNAs for several reasons: first, the considerable cost of chemically synthesized siRNAs; second, the possibility of enforceable and stable expression of shRNAs; and third, the availability of applications of expression constructs in primary cell types (for example, using retroviruses) and in whole organisms (for example, in mouse). We have developed a system to drive expression of shRNAs by placing them under the control of the human RNA polymerase III U6 small nuclear RNA (snRNA) promoter, which normally controls expression of small RNAs in cells. This system has now been demonstrated to be effective both in *vitro*[2, 7–9] and *in vivo* transiently in mouse[10], stably during hematopoiesis[9] and stably in the generation of transgenic mice[11,12]. The pSHAG-MAGIC2 (pSM2) cloning vector (Fig. 1) is roughly equivalent to pSHAG-MAGIC1 (ref. 13) with a few notable exceptions. First, the new cloning strategy is based on the use of a single oligonucleotide that contains the hairpin and common 5′ and 3′ ends as a PCR template (Fig. 2). That is, the oligonucleotide itself serves as template and is amplified by PCR using universal primers that contain an *Xho*I site (within the 5 primer) and an *Eco*RI site (within the 3′ primer) to facilitate cloning (Fig. 3). The resulting PCR fragments are then cloned into the hairpin cloning site of the vector pSM2, the clones are verified by sequencing, and the construct is introduced into the appropriate cell lines, where expression of the miR-30–styled hairpins is driven by the human U6 promoter.

MATERIALS

REAGENTS

Amplification buffer (NEB; provided with Vent DNA polymerase) and Vent DNA polymerase (NEB)
GC-melt PCR reagent (Clontech; provided with Advantage-GC PCR kit)
Dimethyl sulfoxide (DMSO; Sigma)
dNTP solution (10 mM)
Primer oligonucleotides (the sequences shown in Fig. 3) and Template oligonucleotide (as described in step 1; example in **Fig. 2**)
Phenol/chloroform , 1:1 (vol/vol), and 10% chloroform
0.3 M sodium acetate (pH 4.8)
70% (vol/vol) ethanol
Glycogen carrier (Boehringer, Mannheim)
High-melting-temperature agarose (such as MetaPhor Agarose, Cambrex)
Glass bead kit (such as QiaexII gel extraction kit, Qiagen)
Restriction endonucleases *Eco*RI and *Xho*I
PirPlus (PIR1)-competent bacteria (Open Biosystems) and selective antibiotics (chloramphenicol and kanamycin)
T4 DNA ligase (NEB, high-concentration ligase)
Vector: pSM2 (pShag Magic Version 2.0; Open Biosystems)

EQUIPMENT

Thermal cycler programmed with the desired amplification protocol

Design of the template oligonucleotide

PROCEDURE

1| To generate the hairpin primer, select a 'sense' sequence(s) of 22 nucleotides (nt) in length from the coding sequence of the gene of interest for each clone to be constructed. *Only coding sequences are targeted, and each shRNA should be chosen such that it contains more than three mismatches to any other gene. Known single-nucleotide polymorphisms should be avoided, as should common exons targeted in alternatively spliced mRNAs. It is recommended that you create at least 3–6 distinct shRNA clones for each gene to be studied. The link to the hairpin generation program 'RNAi oligo retriever' can be found at www.cshl.org/public/SCIENCE/hannon.html. Either accession numbers from GenBank or raw sequences or siRNA sequences can be used to generate hairpin PCR primers.*

Figure 1 | Retroviral backbone of pSHAG-MAGIC2. The shRNA expression cassette is carried in a validated murine stem cell virus (MSCV) backbone. The 5′ and 3′ flanks are derived from 125 bases of sequence surrounding the human miR-30 miRNA. The pSM2 vector is a redesign of pSM1 that is more stable and less prone to recombination. This vector can be used both for transient delivery by transfection and for stable delivery using the replication-deficient retrovirus as a delivery method. The complete vector sequence is available at www.cshl.org/public/SCIENCE/hannon.html and www.openbiosystems.com.

2| Complete the design of the template oligonucleotide by incorporating miR-30 microRNA (miRNA) sequences (as DNA) into the target sequence (see **Fig. 2** for guidance). Insert the miR-30 loop sequence between the sense and the antisense sequences and add the appropriate flanking sequences of miR-30 miRNA to the 5′ end and the 3′ end of the target. *Figure 2 shows the configuration of a sample template DNA oligonucleotide derived from the RNA oligonucleotide design.* **▲CRITICAL STEP**

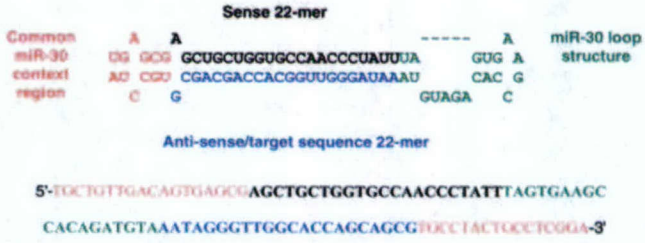


Figure 2 | Design and sequence of a sample template. Above, the final, preprocessed structure of the miR-30–styled shRNA, with derivations of each part of the sequence. Below, the corresponding oligonucleotide DNA template sequence design derived from the oligonucleotide RNA. The sequences in red represents the flanking miR-30 sequences and those in green the miR-30 loop structure. The sample sense- and antisense-selected target sequences are shown in black and in blue, respectively. The mir-30–styled shRNA is synthesized as a single-stranded DNA oligonucleotide with common ends corresponding to part of the endogenous miR-30 miRNA flanking sequence. The flanking regions, shown in red, are used as universal flanks to prime a reaction, whereby the entire miR-30–styled shRNA is amplified to produce a PCR product that may be cloned into pSM2.

3| Synthesize or order the template oligonucleotide designed in steps 1 and 2 and the 5′ and 3′ primer oligonucleotides (Fig. 3). *Because very little primer is required for the PCR reaction, these can be ordered at 0.5 μmol scale from, for example, Sigma-Genosys. We find purification of the oligonucleotides by polyacrylamide gel electrophoresis to be costly and unnecessary.*

Amplification of the template oligonucleotide

4| Set up the amplification reaction (one for each short hairpin sequence to be tested).

For each 100-μl reaction:

10× amplification buffer	10 μl	3′ primer (50 μM stock)	1 μl
DMSO	5 μl	Template oligonucleotide (100 ng/μl)	1 μl
GC melt PCR reagent	5 μl	Vent DNA polymerase	1 μl
dNTPs (10 mM stock)	2 μl	Water	74 μl
5′ primer (50 μM stock)	1 μl		

Note that the GC-melt PCR reagent is available from Clontech only in the Advantage-GC PCR kit, not as an individual reagent. This PCR kit, however, can be used for individual miR-30 shRNA cloning experiments with reasonable success.

5| Amplify the nucleic acids according to the following program:

Cycle number	Denaturation	Annealing	Polymerization	Final
1	1 min at 94 °C	30 s at 54 °C	1 min at 75 °C	
2–24	30 s at 94 °C	30 s at 54 °C	1 min at 75 °C	
Last	30 s at 94 °C	30 s at 54 °C	11 min at 75 °C	
Hold				4 °C thereafter

Times and temperatures may need to be adapted to suit the particular reaction conditions.
➨TROUBLESHOOTING

6| Analyze an aliquot of the amplification reaction by electrophoresis through a polyacrylamide gel, and estimate the concentration and yield of the amplified target. The reaction should produce a single band of 137 bp.

7| To purify the amplification product, extract the reaction once with an equal volume of phenol/chloroform and once with an equal volume of chloroform.
If using a column-based PCR cleanup kit, make sure that the procedure is designed to capture products as small as 100 bp.

5'miR30PCRxhoIF (5' primer):
5'-CAGAAGGCTCGAGAAGGTATATTGCTGTTGACAGTGAGCG-3'

3'miR30PCREcoRIF (3' primer):
5'-CTAAAGTAGCCCCTTGAATTCCGAGGCAGTAGGCA-3'

Figure 3 | Primer oligonucleotides for amplification. The universal PCR primers shown here contain *Xho*I and *Eco*RI cloning sites that allow cloning into *Xho*I and *Eco*RI sites in the pSM2 miR-30 cloning cassette. The portions of sequences in red represent the regions that are complementary to the target shown in **Figure 2**.

8| Recover the DNA by precipitation with ethanol in the presence of 0.3 M sodium acetate (pH 4.8) and wash the pellet in 70% ethanol.
We recommend adding glycogen (at 50 μg/ml final concentration) carrier as a pellet marker.

Cloning of the amplification product into pSM2

9| Dissolve the pellet in 10 mM Tris (pH 7.6) or water and digest the amplification product with *Eco*RI and *Xho*I at 15–25 °C for 90 min.

10| Purify the resulting fragment by electrophoresis through 2% high-melting-temperature agarose, recover the DNA from the gel using the QiaexII glass bead gel purification kit, and resuspend the DNA in 20 μl of 10 mM Tris (pH 7.8). The resulting fragment will be ~114 bp.
Metaphor agarose gives excellent resolution for these separations but is not essential.

11| Prepare the vector DNA by digesting 2 μg of pSM2 with *Eco*RI and *Xho*I, gel purify the linearized plasmid and resuspend it in 20 μl of 10 mM Tris (pH 7.8).

12| Set up the ligation reaction, including 3–5 μl of PCR product, 1 μl of the vector, and T4 DNA ligase, in a 20-μl reaction. Incubate the reaction at 15–25 °C for 10 min to 2 h, as appropriate.

13| Transform 5–10 μl of the ligation reaction directly into PIR1-competent bacteria and select for growth in the presence of both chloramphenicol and kanamycin (each at a concentration of 25 μg/ml).
➨TROUBLESHOOTING
▲CRITICAL STEP

Sequencing and transformation of the cloned shRNA

14| Select and verify by sequencing at least six Cm^R/Km^R bacterial clones for each individual target oligonucleotide when cloning individual shRNAs, or ensure at least threefold coverage when sequencing complex cloning pools (that is, 3–10,000 shRNAs).
The sequence of the U6 sequencing primer (beginning at –42 from the start of transcription) is 5'-GTAACTTGAAAGTATTTCG-3'.

As a matter of practice, we only use sequence-verified shRNAs for biological experiments. We find that the quality of oligonucleotide synthesis can vary considerably with regard to errors introduced during synthesis.

➡TROUBLESHOOTING

15| Transform the verified clones into the cell lines of interest to test their ability to knock down gene expression. The vector(s) may be transiently transfected into any commonly used cell line that also expresses the gene of interest, or transiently transfected with the target gene into a cell line that does not express the gene of interest.

16| Monitor expression of the target gene in the transfected cells by western or northern blotting or by reverse transcriptase-PCR (RT-PCR).

▲CRITICAL STEP

17| Identify the clones that produce the desired knockdown effect and proceed with stable expression experiments to further analyze loss of gene expression.

▲CRITICAL STEP

TROUBLESHOOTING TABLE

PROBLEM	SOLUTION
Steps 5 and 14 The sequences of selected clones do not always match the sequence of the template used in amplification and cloning.	The biggest problem that we have encountered with this protocol is the error frequency associated with PCR amplification. As it turns out, the oligonucleotides used as PCR templates are predicted to fold back on themselves as single-stranded DNA hairpins and have relatively high melting temperatures, often >75 °C. Therefore, we do not recommend deviating from the protocol described above. We have now successfully cloned and sequenced over 45,000 miR-30–styled hairpins using this specific method.
Step 13 We sometimes obtain a high background upon transformation of the ligation reaction products into PIR1-competent cells.	The pSHAG-MAGIC vectors were designed with two bacterial drug selection markers, one on the backbone of the vector (Km^R) and one situated next to the hairpin expression cassette (Cm^R). To minimize background and ensure a successful short hairpin activated gene silencing (SHAG), it is paramount that both markers be used in transformation of PirPlus (PIR1) cells.

▲CRITICAL STEPS

Step 2 The flanking regions of the template (**Fig. 2**; shown in red) are used as universal flanks for PCR amplification. These 5′ and 3′ ~125-nt flanks of the endogenous human miR-30 miRNA were modified to contain *Xho*I and *Eco*RI sites, respectively (see www.open.biosystems for map and sequence). The universal PCR primers (**Fig. 3**) contain *Xho*I and *Eco*RI cloning sites that allow cloning into *Xho*I and *Eco*RI sites in the pSM2 miR-30 cloning cassette. The modified miR-30 context (**Fig. 2**) allows for efficient processing of both the wild-type miR-30 miRNA sequence as well as miR-30–styled shRNAs (data not shown; see also ref. 14).

Step 13 Ligation mixtures must be transformed into PIR1-competent bacteria. The pSM2 plasmid harbors a conditional bacterial origin of replication that requires expression of the *pir1* gene to be rendered functional.

Step 16 For small-scale applications using shRNAs (for example, knocking out a handful of genes), we strongly recommend constructing 3–6 shRNAs per gene and carrying out a validation step before doing the actual biological experiment. For example, we often will transiently transfect shRNA plasmids into any commonly used cell line that also expresses the gene of interest and assay efficacy of the shRNA by western blotting, northern blotting or RT-PCR. Alternately, the target gene can be introduced transiently along with the shRNA and assayed in a cell type lacking endogenous expression. Western blots or RT-PCR are good predictors of efficacy in these assays. We find a direct correlation between whether an shRNA works well in transient assays and whether it works well when expressed stably (for example, from a

retrovirus). In fact, we have been able to build an 'epiallelic' series by this method: a series of shRNAs with different efficacies that give rise to phenotypes of corresponding severity. The correlation of knockdown and phenotype appears to hold true from transient and stable experiments *in vitro* to stable expression *in vivo*[9].

Step 17 One particular plasmid design will not suit all needs. The pSM2 vector therefore is constructed so as to permit high-throughput, automated transfer of the inserted shRNAs between different vector backbones. This feature will permit, for example, rapid migration of the library into other delivery systems, such as lentiviruses and adenoviruses. It will also permit the design of customized delivery vehicles suitable to each screening application.

COMMENTS

An interesting difference with RNAi in mammals is the apparent lack of amplification and transport of the silencing trigger. In contrast to what occurs in *Caenorhabditis elegans* or in plants, mammalian RNAi appears to produce a transient, cell-autonomous targeting effect. These triggers of RNAi in mammalian cells have half-lives and, at least for the time being, must be fed continuously to the RNAi machinery to maintain silencing. Because stable RNAi requires stable expression of the dsRNA trigger, we and others have developed retroviral expression constructs that promote the stable expression of shRNAs[2,9,12,15]. In the particular retroviral vector described in this protocol (pSM2), the shRNA cassette is placed between the 5′ long terminal repeat (LTR) and the drug-selection markers (**Fig. 1**), and the design of the hairpin cassette incorporates sequences of the human miR-30 miRNA (**Fig. 2**). First, adding the miR-30 loop and 125 nt of miR-30 flanking sequence on either side of the hairpin results in more than tenfold greater Drosha and Dicer processing of the expressed hairpins, as compared to the old designs (P.J.P., D. Siolas, A.M. Denli and G.J.H., unpublished observations; see refs. 14,16,17). Increased Drosha and Dicer processing translates into greater siRNA production and greater potency for expressed hairpins. Second, by using the miRNA-30 designs, we can incorporate 'rule-based' designs for target sequence selection. One such rule is the destabilization of the 5′ end of the antisense strand, which results in strand-specific incorporation of miRNAs into RISC[18]. Last, the miR-30 design offers more flexibility in applications of gene silencing, as these sequences can also be expressed from RNA polymerase II promoters (for example, cytomegalovirus) or even arrays of different hairpins in polycistronic transcripts (data not shown; see ref. 19).

SOURCE

G.J. Hannon is the editor of *RNAi: A Guide to Gene Silencing* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2003). This protocol is a new development of ideas discussed in that volume.

1. Hannon, G.J. RNA interference. *Nature* **418**, 244–251 (2002).
2. Paddison, P.J. & Hannon, G.J. siRNAs and shRNA: skeleton keys to the human genome. *Curr. Opin. Mol. Ther.* **5**, 217–224 (2003).
3. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
4. Caplen, N.J., Parrish, S., Imani, F., Fire, A. & Morgan, R.A. Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl. Acad. Sci.* **98**, 9742–9747 (2001).
5. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
6. Paddison, P.J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
7. Paddison, P.J., Caudy, A.A. & Hannon, G.J. Stable suppression of gene expression by RNAi in mammalian cells *Proc. Natl. Acad. Sci.* **99**, 1443–1448 (2002).
8. Paddison, P., Caudy A.A., Bernstein, E., Hannon G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
9. McCaffrey, A.P. *et al.* RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
10. Hemann, M.T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nat. Genet.* **33**, 396–400 (2003).
11. Carmell, M.A., Zhang, L., Conklin, D.S., Hannon G.J. & Rosenquist, T.A. Germline transmission of RNAi in mice. *Nat. Struct. Biol.* **10**, 91–92 (2003).
12. Rubinson, D.A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat. Genet.* **33**, 401–406 (2003).
13. Paddison, P.J., Caudy, A.A., Sachidanandam, R. & Hannon, G.J. Short hairpin activated gene silencing in mammalian cell. *Methods Mol. Biol.* **265**, 85–100 (2004).
14. Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
15. Brummelkamp, T.R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247(2002).
16. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
17. Zeng, Y. & Cullen, B.R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
18. Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
19. Zeng, Y., Wagner, E.J. & Cullen, B.R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell.* **9**, 1327–1333 (2002).

Oncogene (2004) 23, 8401–8409

www.nature.com/onc

RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age

Jose Silva[1], Kenneth Chang[1], Gregory J Hannon*[1] and Fabiola V Rivas[1]

[1]*Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA*

Sequencing of complete genomes has provided researchers with a wealth of information to study genome organization, genetic instability, and polymorphisms, as well as a knowledge of all potentially expressed genes. The identification of all genes encoded in the human genome opens the door for large-scale systematic gene silencing using small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs). With the recent development of siRNA and shRNA expression libraries, the application of RNAi technology to assign function to cancer genes and to delineate molecular pathways in which these genes affect in normal and transformed cells, will contribute significantly to the knowledge necessary to develop new and also improve existing cancer therapy.
Oncogene (2004) **23,** 8401–8409. doi:10.1038/sj.onc.1208176

Keywords: RNAi; high throughput screening; functional genomics; cancer; apoptosis; synthetic lethality

Introduction

RNA interference (RNAi) is a conserved biological response discovered in the nematode *Caenorhabditis elegans*, as a response to double-stranded RNA (dsRNA). Initially demonstrated by Mello and coworkers, who showed that injection of long dsRNA into *C. elegans* led to sequence-specific degradation of the corresponding mRNAs, this silencing response has been subsequently found in other eukaryotes from yeast (*Neurospora crassa* and *Schizosaccharomyces pombe*) to mammals (Fire *et al*., 1998; Hannon, 2002; Montgomery, 2004). Although knowledge of the biological mechanism of RNAi has grown exponentially over the last few years, application of RNAi at the genome-wide level had to await the development of optimal techniques of delivery. These were pioneered in model organisms; for example, RNAi can be triggered by soaking *C. elegans* (Tabara *et al*., 1998) and *Drosophila* cells (Clemens *et al*., 2000) in a solution of dsRNA, or by feeding worms with *Escherichia coli* expressing gene-specific dsRNAs (Timmons and Fire, 1998). In mammalian cells, however, long dsRNAs (>30 nucleotides) elicit an antiviral interferon response (Minks *et al*., 1979; Manche *et al*., 1992). Thus, RNAi technology could not be applied to mammals until the discovery that short dsRNA duplexes, processed from long dsRNA into 21–28 cleavage fragments termed small interfering RNAs (siRNAs), were sufficient to trigger gene-specific silencing upon transfection into mammalian cells (Elbashir *et al*., 2001; Harborth *et al*., 2001). However, the silencing response to transfected siRNAs is transient, lasting from 3 to 7 days depending upon the rate of cell division making this approach unsuitable for analysis of long-term effects of silencing. The search for a more sustained silencing response has resulted in the development of an additional class of triggers, short hairpin RNAs (shRNAs), that can establish stable gene silencing by continuously supplying the RNAi trigger (see for review Paddison and Hannon, 2002). Researchers are now using this technology to understand biological mechanisms in both normal cells and in malignant ones, one of the major goals being to unravel the mysteries of transformation and to improve current cancer therapy.

Viewing cancer as a global epidemic of discrete afflictions is a confounding oversimplification. Each cancer is a unique disease arising from multiple genetic alterations, and the particular combination of genes mutated in any given patient probably determines the degree of malignancy and potential therapeutic vulnerabilities of that individual's cancer. Improved prevention, diagnosis, and treatment of cancer in patients, will require a detailed understanding of the specific molecular mechanisms that go away in specific cancers. This understanding must be derived both from an examination of the cancerous cell itself and from an investigation of the interactions between the cancer and its host. One of the ways in which these insights can be obtained is through functional genetic approaches in mammals.

Traditionally, functional genetic studies are divided into forward or reverse screens. In a typical forward genetic study, genes are mutated at random. The resulting changes in the phenotype of a cell or organism are then attributed to the mutated genes and, by inference, to their protein products. After identification of an abnormal phenotype, the mutations must be mapped, a process which is usually time-consuming and not easily applicable to mammalian systems. Conversely, reverse genetic approaches involve the disruption of a gene of interest, so as to determine its function and/

*Correspondence: GJ Hannon, Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA;
E-mail: hannon@cshl.edu

or involvement in a pathway. Classical reverse genetic approaches involve the creation of knockout cell lines or organisms, and can be expensive, time-consuming and unsuitable for genome-wide screens in mammals. Faster, simpler and cheaper alternatives of attenuating gene function in a sequence-specific manner have emerged in the form of antisense technology, ribozymes and more recently, RNAi (Figure 1).

To facilitate the use of RNAi in mammals, several groups have constructed first-generation RNAi libraries of shRNA expression vectors. For example, we have constructed a library comprising approximately 28 000 shRNA expression cassettes targeting 9610 human and 5563 mouse genes. The expression cassettes in our collections are sequence-verified and contained within multifunctional vectors that can be packaged into retroviruses, tracked in a mixed cell population by means of a random 60-mer DNA 'barcode', and shuttled into customized vectors through bacterial mating (Paddison *et al*., 2004). Bernards and colleagues have constructed a similar arrayed library with overlapping functionality (Berns *et al*., 2004). Given advances in our understanding of the RNAi mechanism over the last year, we have also constructed a second-generation library that uses improved expression cassettes and informatic tools for shRNA design. This library is presently available to academic investigators through several sources (e.g., Open Biosystems, MRC Geneservice). Given the availability of these powerful tools, there is a pressing need to discuss possible screening methodologies that are available for mammalian systems and how such approaches compare to those available in more traditional genetic models. Such a discussion is the purpose of this brief review.

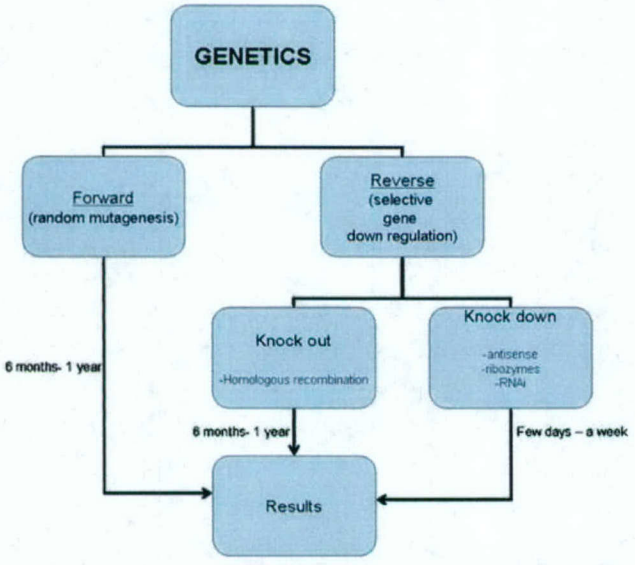


Figure 1 The diagram shows a comparison between forward and reverse genetics. A typical forward genetic study begins with the generation of random mutants in order to produce a specific phenotype. The next step is the identification of the genetic alteration that originated the phenotype. In mammalian cells, the time required to complete the study could be estimated between 6 months–1 year. Classical reverse genetic involves the selective downregulation of a gene function. Although knockout techniques produce the total abrogation of a protein by disruption of the two alleles of the genome, the time necessary to complete the process is also very long (6 months–1 year). On the other hand, new approaches that can achieve 90–95% of suppression of the gene expression like antisense, ribozymes or more recently RNAi are more simple and a much faster alternative

Genetic screens in mammalian cells

The ability to perform genetic screens has been popularized over the last 30 years primarily through the use of model organisms such as bacteria (Shuman and Silhavy, 2003), yeast (Forsburg, 2001), worms (Jorgensen and Mango, 2002), and flies (St Johnston, 2002). The demand for better tools to assign gene function has been made imperative by the advent of genomics, a field that within a few years has produced numerous monumental advances from the complete sequence of *S. cerevisiae* (Dujon *et al*., 2004) to a draft sequence of the human genome. (Celera Genomics Project, 2001; International Human Genome Sequencing Consortium, 2001). The availability of such copious sequence information has thrown into sharp relief the need for versatile technologies for decoding gene function. Nowhere has this been more apparent than in mammalian systems where only a few years ago the existing approaches to functional genomics offered few options. There have been isolated successes with overexpression screens (Michiels *et al*., 2002; Huang *et al*., 2004), insertional mutagenesis (Mikkers and Berns, 2003) and genome-wide two-hybrid studies (Chen and Han, 2000). However, approaches for routine, loss-of-function genetics on a large scale, particularly in cultured cells, were lacking.

In principle, cultured mammalian cells have many of the benefits associated with yeast as models for the study of eukaryotic cell genetics. Like yeast, they are amenable to gene transfer in mass or individually, a variety of selectable/detectable markers is available for establishing expression of heterologous genes, constructs can be maintained extrachromosomally (transient) or stably integrated via viral vectors, and in almost all cases, different genetic lines can be established from mammalian cell lines by cloning at limiting dilution. Additionally, mammalian cell lines present important advantages to yeast as models not only of cell genetics of higher eukaryotes, but also of biological process that have an impact at the organismal level. Accumulating data show that many human and murine genes are not represented in yeast, and there are no true counterparts to cellular processes such as transformation, apoptosis, tissue-specific differentiation and some signaling pathways in these lower eukaryotes (Aravind *et al*., 2001).

The technological shortfall in mammalian functional genomics has been, to a degree, met by recent advances toward the creation of routine RNAi-based tools for gene silencing in cultured cells and in animals. The initial finding that siRNAs and encoded shRNAs could trigger gene silencing in mammals has been extended by

the combination of RNAi with viral and episomal vectors that allow stable maintenance of silencing in mammals (Paddison and Hannon, 2003). This has reinvigorated and broadened interest in using mammalian cells for both forward and reverse genetic screens.

There are a number of possible approaches to screening large collections of silencing triggers for their effects on mammalian cells, and several of those will be discussed herein as examples of how RNAi libraries may be applied to cancer-relevant biological problems in mammalian cells. Three screening modalities will be discussed. First, individual siRNAs or shRNAs can be transfected and screened in multiwell format for activation or repression, of a reporter or activity in a cell-based or biochemical assay (Somma *et al*., 2002; Aza-Blanc *et al*., 2003; Brummelkamp *et al*., 2003; Boutros *et al*., 2004; Hsieh *et al*., 2004; Paddison *et al*., 2004). In this format, individual genes are transiently suppressed 'one-by-one' and analysis is carried out in a high throughput manner using a robotic platform. Second, cells can be infected with pools of shRNAs (at MOI of 1.0), followed by selection of individual colonies that can be scored phenotypically for alterations that result from the expression of a specific trigger (e.g., a morphological alteration). A variant of this procedure allows siRNAs or shRNAs to be delivered *in situ* via 'reverse transfection' with resulting phenotypes being scored in transfected microcolonies. Third, transduced pools of cells can be monitored in mass, via DNA 'barcodes' contained within the shRNAs, by high-density oligonucleotide microarrays for relative changes in shRNA representation following application of a selective stimulus (Paddison *et al*., 2004; Berns *et al*., 2004). A variant of these methods involves long-term selection for growth of colonies under specific conditions. This is defined as a genetic selection rather than a screen and will be discussed separately, below. Of course, the distinction between genetic selection and genetic screen can be subject to interpretation, particularly as modern techniques (e.g. barcode arrays) are mapped onto these longstanding definitions.

Transient screens (cell-based assays)

In these types of assays, cells are plated for high throughput studies, normally in 96- or 384-well plates, and each individual well is transfected with a different siRNA, shRNA or limited-complexity pool of triggers (Figure 2). The transfection of the RNAi triggers is usually mediated by lipid-based reagents that allow for high and reproducible transfection efficiencies on robotic platforms. An enormous variety of lipid-based transfection agents are already available in the market, and allow for transfection of siRNAs, plasmids, or both.

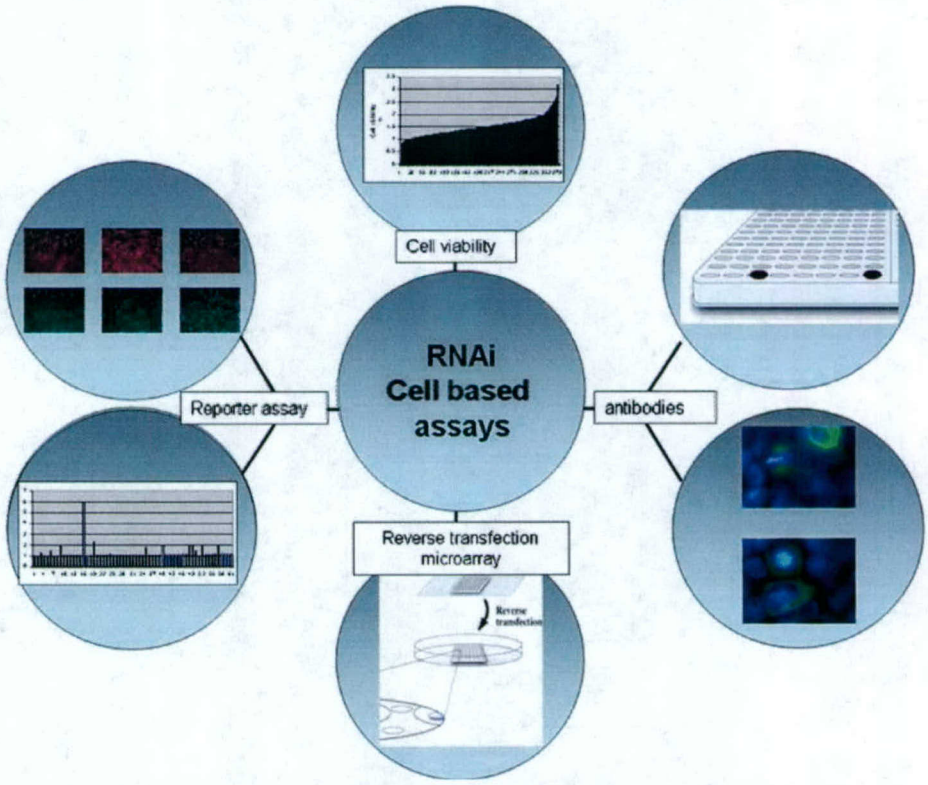


Figure 2 The graphic shows different possibilities for performing cell-based screens using RNAi. When combined with a simple cell viability assay it can identify essential genes. Fluorescent and luminescent reporters that are activated by a specific pathway offer the possibility to identify more detailed phenotypes. Additionally, classic immunofluorescence techniques increase the number of phenotypes that can be studied. The use of automated microscopy platforms allows the realization of high through put screens that can identify even light morphological abnormalities. Finally, the adaptation of these screens to a miniaturized format will simplify enormously the realization of these studies

Selection of the transfection carrier depends on both the cell type and the RNAi silencer being used, and their efficiency should be tested empirically with both positive and negative controls. In addition to classical methods of transfection, a novel approach named 'reverse transfection' has been recently described (Ziauddin and Sabatini, 2001). This technique is especially suitable for high throughput screens. A mixture of transfection reagent and nucleic acids is dispensed on a well of a tissue culture plate and dried for storage. The cells plated on the transfection layer incorporate the lipid–nucleic acid complexes, and after several hours, the effect of the transfection can be examined.

A promising new technology that greatly expands the potential throughput of RNAi-based screens adapts the reverse transfection method to a microarray-based platform. Mammalian cells are plated on a glass slide spotted in defined locations with transfection mixtures containing different RNAi triggers (Kumar *et al.*, 2003; Mousses *et al.*, 2003; Silva *et al.*, 2004). Cells growing on the printed areas take up the nucleic acids, creating spots of localized transfection within a lawn of nontransfected cells. Although theoretically this technology shows a great potential for genome-wide analysis (thousand of knockdowns can be analysed on a slide), there are still technical limitations to be overcome. For instance, because of the small number of cells that are transfected (between 50 to a few hundred cells), variability and sensitivity are parameters that may compromise the results. Moreover, studies that require long incubation times to reveal a phenotype may not be amenable to this approach. Most importantly, there are thus far a limited number of cell lines that have been reported to have transfection efficiencies high enough to be used in this procedure, thus limiting its application.

Once the approach for performing an RNAi screen has been chosen and optimized, developing a phenotypic assay is the next step. There are an almost unlimited number of strategies to identify the phenotype of interest. Initial genome-wide screens have examined a simple phenotype: cell proliferation versus cell death. For instance, Aza-Blanc *et al.* (2003) described the application of an RNAi-based genetic screen in mammalian cells toward understanding the biology and mechanism of TRAIL-induced apoptosis. TRAIL is a TNF superfamily member that induces cytotoxicity in tumor cells when bound to its cognate receptors. Binding of TRAIL to specific death receptors (DR4 and DR5) induces apoptosis through recruitment of adaptor molecules, which results in the formation of the death-inducing signaling complex and the activation of downstream apoptotic pathways. To identify genes that modify cellular sensitivity to TRAIL-induced death, the authors screened a limited complexity siRNA library using HeLa cells in the presence or absence of TRAIL. After an incubation period, cell viability was measured by addition of a dye (Alamar Blue) that produces a fluorescent signal, which is reflective of the extent of cellular proliferation. Their screen, validated by the identification of known apoptotic and anti-apoptotic genes, also led to a functional linkage of genes like *DOBI*, *MIRSA*, GSK3α or *SRP72* to the TRAIL-mediated response. Although this study was limited to several hundred genes, it illustrates the potential of a genome-wide RNAi analysis in mammals aimed at identifying genes involved in cell growth and viability. A similar study has been carried out in Drosophila cells by Boutros *et al.* (2004), who identified 438 essential genes by simple quantification of cell number after transfection of individual dsRNAs to each of Drosophila's predicted 20 000 genes. Analogous assays can be designed by measuring parameters such as caspase activation, ATP content, or cell membrane permeability. In all cases mentioned, the effect of the RNAi trigger must extend to the vast majority of cells, so that loss of viability in a subpopulation is not masked by unaffected cells. This limits the approach to cell lines that are very easily transfected under the chosen cell culture conditions. The great advantage of this assay is its simplicity. However, such approaches are not illuminating regarding precisely which biological pathway is affected to elicit cell death.

More informative assays require that more specific phenotypes be examined. We and others have explored the use of reporter systems in which the activity of a protein or cellular pathway is monitored by easily detectable changes in expression (or activity) of a transgene (luciferase, fluorescent proteins, CAT, β-gal, etc...). For example, Brummelkamp *et al.* (2003) used this strategy to study the family of ubiquitin-specific proteases (deubiquinating enzymes or DUBs). Post-translational modification by conjugation of ubiquitin moieties plays a major role in the control of protein half-life and thus in their activity. Ubiquitin conjugating enzymes and DUBs mediate ubiquitination and deubiquitination, respectively, of cellular substrates. These families maintain the correct balance between how much protein is driven to degradation and how much protein is preserved. The authors designed a collection of RNA interference vectors to suppress 50 human DUBs, and searched for those relevant to the NF-κB pathway. They cotransfected an NF-κB-luciferase reporter gene and different DUB knockdown vectors into human cells, and measured the effect of DUB knockdown on tumor necrosis factor-α- (TNF-α) mediated activation of NF-κB. Importantly, they found that RNAi targeting of the cylindromatosis tumor suppressor gene *CYLD* enhanced the activation of an NF-κB reporter. Similarly, we have used a reporter approach to detect changes in the activity of the multisubunit 26S proteosome, the major nonlysosomal protease in eukaryotic cells, by using, as a readout, a green fluorescent protein genetically modified to be a target for degradation by the proteasome (Paddison *et al.*, 2004). We tested an shRNA library of approximately 7000 constructs for the ability to block proteasome-mediated proteolysis, as reflected by the accumulation of the modified fluorescent protein. Our study revealed approximately 100 RNAi constructs that increased the accumulation of the reporter out of which 22 corresponded to 15 known proteasome subunits. This screen was conducted by

cotransfection of the reporter and shRNA expression vector into cell plated in 96-well plates. However, we have obtained similar results using a screening protocol in which RNAi was triggered *in situ* by reverse transfection on microarrays (Silva *et al*., 2004).

Although reporter constructs are a very convenient approach for cell-based assays, appropriate reagents are not always easily available to assay interesting phenotypes. In some cases, conventional techniques such as immunofluorescence (IF), have presented a feasible alternative. In a proof of concept experiment, Hsieh *et al*. (2004) used this approach to identify inhibitors of the phosphatidylinositol-3-kinase(PI3K)/Akt signaling pathway. Members of the PI3K family are characterized by their ability to phosphorylate the inositol ring 3′-OH group in inositol phospholipids, generating the second messenger phosphatidylinositol-3,4,5-triphosphate (PI-3,4,5-P(3)). This compound in turn recruits Akt to the inner cell membrane, where the kinase becomes phosphorylated and activated. Activated Akt modulates the function of numerous substrates involved in the regulation of cell survival, cell cycle progression, and cellular growth. Hsieh *et al*. tested functionally a set of siRNAs in a screen aimed at identifying negative regulators of the Akt phosphorylation. In this screen, upon siRNA transfection, modulation of the Akt phosphorylation was detected by IF staining with anti-phospho-Akt (detecting Akt phosphorylation at S473). As expected, the known Akt regulator PTEN scored positive in the screen, validating this approach for cell-based RNAi studies.

Some biological questions cannot be answered by examining changes in the intensity of a reporter or monitoring site-specific phosphorylation of a molecule. Instead, scoring a visible phenotype might be required to assay protein function. Such studies may be well suited for automated microscopy in which high-content images are automatically analysed by sophisticated image processing software to determine phenotype. The study of genes involved in different aspects of cytokinesis presents a clear example of such a process. As a proof-of-principle experiment, we knocked down the mitotic motor protein Eg5, as cytokinesis defects in cells where Eg5 function is inhibited are well established (Silva *et al*., 2004). Transfection mixtures contained a plasmid encoding an α-tubulin GFP fusion protein and individual shRNAs targeting Eg5. In this experiment, the GFP fusion protein identifies the cells that have been transfected, and also allows visualization of microtubules. Microscopic analysis of the transfected cells revealed a 'rosette' pattern characteristic of the cells displaying loss of kinesin Eg5. Similar results were obtained by IF staining with anti-α-tubulin antibodies. Using an analogous approach, Somma *et al*. (2002) present a nice example of molecular dissection of the cytokinesis pathway in *Drosophila* cells. Their phenotypic analysis identified genes required for different aspects of cytokinesis, such as central spindle formation, actin accumulation at the cell equator, contractile ring assembly or disassembly, and membrane behavior.

The aforementioned examples present a few of the limitless possibilities that can be explored by combining RNAi with assay systems that are limited only by the imagination of the investigator. However, many phenotypes require significant time to develop or assay or can only be probed by reintroduction of cells into a tissue setting.

Genetic screens and selections with stable populations

The need for examining phenotypes that develop over a time span that reaches from several days to several weeks has been addressed by the development of methods for stably integrating shRNA expression cassettes into the genome of target cells. In this mode, the shRNA expression library is most commonly packaged into retroviruses, transduced into cells, and stable integrants selected such that each cell is targeted to carry, on average, one copy of the hairpin expression cassette. There are several advantages to this approach over transiently transfected screens: The knockdown effects can be monitored over extended periods, shRNA expression is more normalized, thereby facilitating the screening of cells in pools, and finally, this approach is very adaptable for high throughput studies.

Once a population of cells that stably expresses shRNAs is produced, two alternative approaches can be undertaken for assaying the consequences of gene knockdown. Cells can be plated at low density to ascertain their phenotypic behavior through a positive selection. In this mode, only cells that have a specific characteristic such as the ability to proliferate under specific conditions (e.g. colony formation in soft agar, focus formation, insensitivity to growth inhibitory cytokines, etc.) will be selected. Following selection, individual colonies can be isolated and the identities of integrated shRNA cassettes determined by sequencing. In an alternative approach, the fate of shRNAs in a population of virally transduced cells can be monitored by adopting a DNA-barcoding strategy, which has been previously used in *S. cerevisiae* for following complex populations of mutants by DNA microarrays (Shoemaker *et al*., 1996).

The former approach has been well validated using both integrated and episomal cDNA libraries that cause ectopic expression of certain genes and with antisense RNA libraries that can inhibit gene expression. With shRNA libraries, a genetic selection was used by Berns and colleagues to investigate components of the p53 growth arrest pathway. These investigators used conditionally immortalized primary human fibroblasts that undergo a senescence program upon reassertion of the p53 and Rb tumor suppressor pathways. These cells were screened using a library of 23 703 shRNAs targeting 7914 different human genes (Berns *et al*., 2004) for constructs that allowed continued proliferation in face of reactivation of tumor suppressor function. Using colony formation assays, and screening 83 different populations of transduced cells, they identified six genes, all in the p53 pathway and including *p53* itself, which when suppressed, conferred resistance to both *p53*-dependent and *p19^{ARF}*-dependent prolifera-

tion arrest and also abolished DNA damage-induced G1 arrest by ionizing radiation.

Aside from successful applications in yeast and phage, the barcode strategy has not been formally validated as a mechanism to identify genes linked to specific phenotypes in mammalian cells. However, the potential power of this strategy has made it a tantalizing and irresistible possibility for applying complex RNAi libraries. There are several approaches to the use of molecular barcodes for tracking complex populations of RNAi triggers, and two have been exemplified in recent publications describing shRNA libraries. One possibility is to use the shRNA sequence itself as a barcode. Another is to link the shRNA sequence to an independent unique sequence within each vector. Irrespective of the barcoding strategy, the underlying concept is that by PCR amplification of integrated DNAs, one may essentially count (relatively speaking) the number of cells that contain a specific shRNA cassette. This is measured by hybridizing genomic PCR products containing the barcodes to custom microarrays that contain the complement of these sequences. By comparing barcode representations from cell populations treated in different ways, one may simultaneously assess the consequences of expressing a given individual shRNA on the cell's response to the treatment (Figure 3). If, for example, a particular shRNA provided resistance to a growth inhibitor stimulus, then the representation of its associated barcode should be increased after treatment. If a given shRNA sensitized a population to a specific stress, then the relative abundance if its barcode should diminish after the stress.

Despite the enormous promise of this approach, there are certain considerations that must be taken into

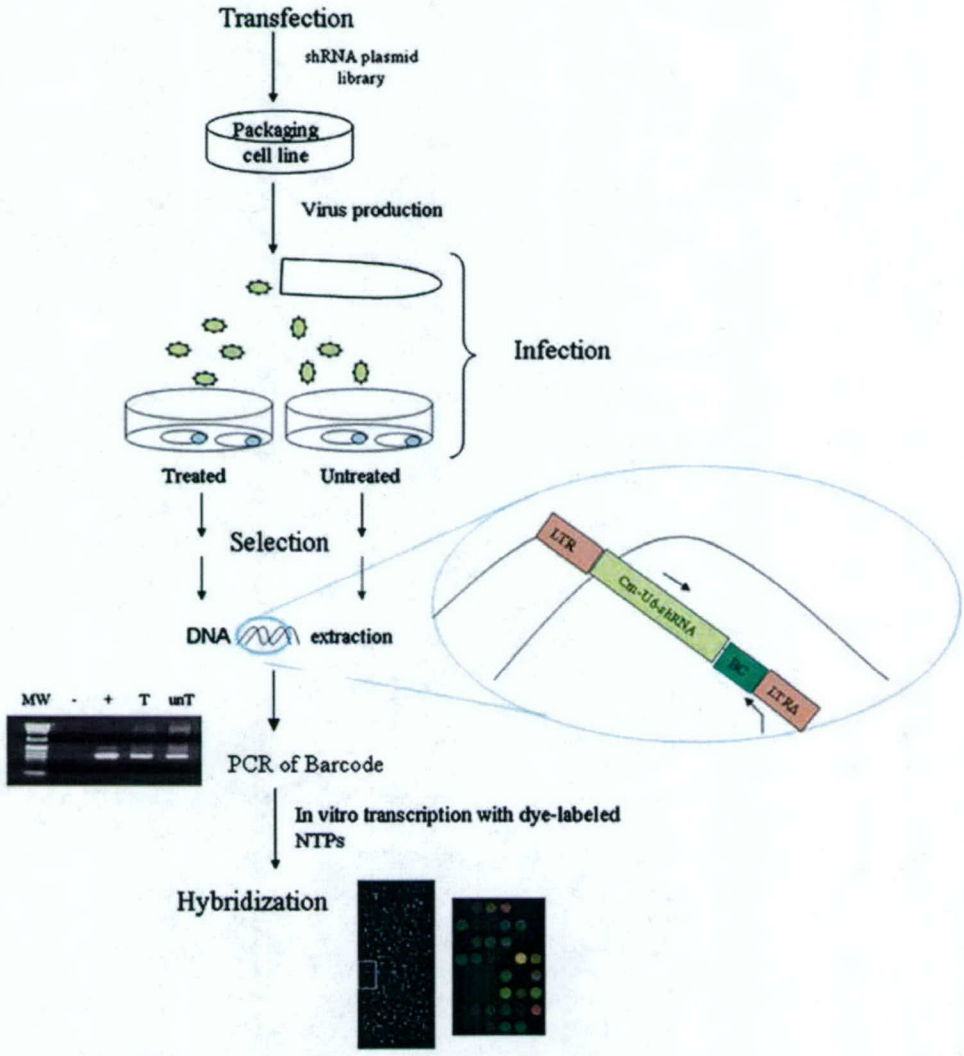


Figure 3 Sample protocol that illustrates a typical RNAi screen in a stably selected population. ShRNAs libraries are packaged into viruses, which are used to infect a target cell population. The populations are then divided into treatment groups, which are subjected to selection according to the experimental setting. After the desired time, DNA is extracted from each sample group and assayed for the presence of DNA barcodes by PCR. One of the PCR primers harbors a T7 RNA polymerase promoter sequence that allows for *in vitro* transcription of single-stranded RNA that is subsequently hybridized to an array containing the complement of the barcode sequences in the entire shRNA library. Comparison between the hybridization patterns of different treatment groups allows for the identification of shRNAs represented in each selected group

account when designing a barcode-based RNAi screen. The first is the potential noise associated with simultaneous analysis of complex populations by any given procedure. The second is the intrinsic noise produced by the chosen biological system.

Noise in the technical procedure used to measure barcode dynamics is easily defined and readily controlled. First, the PCR procedure used to amplify barcodes from the genome of infected populations must accurately reflect the relative proportions of individual members of the population. This is aided by the use of common primers for the amplification of each barcode sequence. Additionally, amplification should be kept in the linear range, which can be easily measured using a modification of a Q-PCR protocol. Remaining variables are applicable to any microarray procedures and include variation in labeling and hybridization. These are discussed in any of a number of published guides to microarray methodologies (for example, Churchill, 2002; Chuaqui *et al*., 2002; Forster *et al*., 2003). One source of potential problems, though not necessarily of noise, is the relative performance of individual probes on the arrays. At present, irrespective of the particular barcode strategy, none of the sequences being used for barcoding is specifically designed as a hybridization probe. While future advances may change this fact, it must be accepted that at present, some probes will simply perform poorly, giving weak signals irrespective of their frequency in the population. Thus far, we have demonstrated the validity of the technical approach to barcode amplification from the genomes of infected cells using populations with complexities reaching 20 000 individual elements.

Noise can also be introduced into the system from the biological components of the assay. Of principal concern in an outgrowth experiment is whether the genetic drift (Glass, 1954) of the population introduces random variation into the relative frequency of each shRNA clone irrespective of its specific biological function. Simulation studies suggest that such drift is most pronounced as complexity rises and population size falls. This leads to each member of the population being represented in relatively few cells. As the population proliferates and is stressed by propagation in culture, certain elements of the population may be lost. For example, most cell lines, split by trypsinization, do not have a 100% plating efficiency. Thus each passage creates a random sample of the total population. If each shRNA is present in only a few cells, there is a relatively high probability that some shRNAs will be lost at each sampling. A second problem is that cell lines are not completely uniform at the level of the individual cell. Thus, each cell within the cell line may have a different probability of doubling for each doubling of the population as a whole. This can also contribute to the loss or amplification of shRNAs over time. We find that both of these problems can be solved to a large degree by insuring that each shRNA in the population is represented in a large number of cells, as is predicted by population genetics simulations (Sachidanamdan *et al*., unpublished).

Given solutions of the aforementioned technical barriers, there are a number of ways in which barcode microarrays can be applied to RNAi-based screens. The simplest is to use barcode arrays rather than analysis of individual colonies to examine populations that have been positively selected following application of a stress. More complicated and fraught with difficulty is the use of barcodes to identify shRNAs that sensitize cells to a treatment or genetic lesion in a so-called synthetic lethal experiment.

Synthetic-lethal screens

A synthetic lethal interaction is defined by a situation in which two nonessential genes become essential when mutated in combination in the same cell (Basson *et al*., 1987; Bender and Pringle, 1991). Such an interaction implies that two genes lie in discrete pathways that normally compensate, each for the loss of the other. When both pathways are lost, death occurs because of a catastrophic loss of a specific biological process. In principle, a similar genetic interaction could also occur when multiple hypomorphic mutations in the same pathway reduce flux through that pathway to levels that are not tolerated by a cell. Carrying out synthetic lethal screens in genetically tractable models such as yeast or *C. elegans* is relatively straightforward. However, the development of this technology in mammalian cells is much less straightforward. In theory, genome-wide RNAi libraries offer a route toward this powerful method for detecting genetic interactions.

The ability to detect cell extinction events on a large-scale is key to realizing the use of RNAi in genome-wide screens for synthetic lethal interactions. There are several ways to accomplish this goal, including the use of highly parallel screens in which the effects of individual constructs are examined in multiwell dishes. In that mode, the successful identification of synthetic lethal interactions depends heavily on a combination of the timing of the event and achieving the appropriate cell plating density to enable measurement of growth inhibition. A more convenient way to approach the problem would be through the use of the DNA barcode strategy to follow loss of cell populations expressing certain shRNAs. In theory, the underlying methodology is straightforward. Cells of two discrete genotypes could be engineered with an shRNA library and then the two populations could be compared following some period of outgrowth. A slightly more sophisticated version would use a single engineered cell population carrying a conditional allele of the target gene. This could be generated either by conventional means or through the introduction into part of the population of a target-specific siRNA. Confounding such approaches is the incomplete expressivity of shRNA-mediated silencing. Even an effective shRNA will not silence its target to the same degree in all cells harboring the shRNA expression cassette. This is due both to position effects on the integrated cassettes and to the propensity to select for cells that have silenced or altered expression cassettes that encode deleterious genes. These problems effectively set a

background that might mitigate against detection of lethal interactions because of an insufficient power in the system to detect relatively minor shifts in populations. As an example, imagine that 70% of cells infected with a given shRNA virus show a synthetic lethality in combination with p53 loss. This would translate into the ability to see only an ~3-fold change in the representation of that shRNA on a barcode microarray, which is probably at the edge of what can be confidently detected. Therefore, our challenge in moving toward the goal of genome-wide synthetic lethal analysis is to increase both the quality of our RNAi reagents and the mechanisms to detect loss of such cell populations.

Conclusions

Over the past several years, we have made tremendous progress in our quest to harness the RNA interference pathway as a method for genetic manipulation of mammalian cells and animals. Last year, with the report of several efforts to construct genome-wide libraries of siRNAs and shRNAs for multiple mammalian models, we have entered a new era in which the application of powerful genetic approaches can be used both *in vivo* and *in vitro* to deepen our understanding of the cancer cell and to identify and validate potential new approaches to cancer therapy.

Acknowledgements
We thank JP O'Keefe for critical reading and suggestions. JS is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research Program. FVR is a Jane Coffin Childs Memorial Cancer Research Fund fellow. GJH is supported by an Innovator Award from the US Army Breast Cancer Research Program and by grants from the National Institutes of Health.

References

Aravind L, Dixit VM and Koonin EV. (2001). *Science*, **291,** 1279–1284.
Aza-Blanc P, Cooper CL, Wagner K, Batalov S, Deveraux QL and Cooke MP. (2003). *Mol. Cell*, **12,** 627–637.
Basson ME, Moore RL, O'Rear J and Rine J. (1987). *Genetics*, **117,** 645–655.
Bender A and Pringle JR. (1991). *Mol. Cell Biol.*, **11,** 1295–1305.
Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL and Bernards R. (2004). *Nature*, **428,** 431–437.
Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, Haas SA, Consortium HF, Paro R and Perrimon N. (2004). *Science*, **303,** 832–835.
Brummelkamp TR, Nijman SM, Dirac AM and Bernards R. (2003). *Nature*, **424,** 797–801.
Celera Genomics Project (2001). *Science*, **291,** 1304–1351.
Chen Z and Han M. (2000). *Bioessays*, **22,** 503–506.
Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V and Emmert-Buck MR. (2002). *Nat. Genet.*, **32** (Suppl), 509–514.
Churchill GA. (2002). *Nat. Genet.*, **32** (Suppl), 490–495.
Clemens JC, Worby CA, Simonson-Leff N, Muda M, Maehama T, Hemmings BA and Dixon JE. (2000). *Proc. Natl. Acad. Sci.*, **97,** 6499–6503.
Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P and Souciet JL. (2004). *Nature*, **430,** 35–44.
Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K and Tuschl T. (2001). *Nature*, **411,** 494–498.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC. (1998). *Nature*, **391,** 806–811.
Forsburg SL. (2001). *Nat. Rev. Genet.*, **2,** 659–668.
Forster T, Roy D and Ghazal P. (2003). *J. Endocrinol.*, **178,** 195–204.
Glass B. (1954). *Adv. Genet.*, **6,** 95–139.
Hannon GJ. (2002). *Nature*, **418,** 244–251.
Harborth J, Elbashir SM, Bechert K, Tuschl T and Weber K. (2001). *J. Cell Sci.*, **114,** 4557–4565.
Hsieh AC, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S and Sellers WR. (2004). *Nucleic Acids Res.*, **32,** 893–901.
Huang Q, Raya A, DeJesus P, Chao SH, Quon KC, Caldwell JS, Chanda SK, Izpisua-Belmonte JC and Schultz PG. (2004). *Proc. Natl. Acad. Sci.*, **101,** 3456–3461.
International Human Genome Sequencing Consortium (2001). *Nature*, **409,** 860–921.
Jorgensen EM and Mango SE. (2002). *Nat. Rev. Genet.*, **3,** 356–369.
Kumar R, Conklin DS and Mittal V. (2003). *Genome Res.*, **13,** 2333–2340.
Manche L, Green SR, Schmedt C and Mathews MB. (1992). *Mol. Cell Biol.*, **12,** 5238–5248.
Michiels F, van Es H, van Rompaey L, Merchiers P, Francken B, Pittois K, van der Schueren J, Brys R, Vandersmissen J, Beirinckx F, Herman S, Dokic K, Klaassen H, Narinx E, Hagers A, Laenen W, Piest I, Pavliska H, Rombout Y, Langemeijer E, Ma L, Schipper C, Raeymaeker MD, Schweicher S, Jans M, van Beeck K, Tsang IR, van de Stolpe O, Tomme P, Arts GJ and Donker J. (2002). *Nat. Biotechnol.*, **20,** 1154–1157.
Mikkers H and Berns A. (2003). *Adv. Cancer Res.*, **88,** 53–99.
Minks MA, West DK, Benvin S and Baglioni C. (1979). *J. Biol. Chem.*, **254,** 10180–10183.
Montgomery MK. (2004). *Methods Mol. Biol.*, **265,** 3–21.
Mousses S, Caplen NJ, Cornelison R, Weaver D, Basik M, Hautaniemi S, Elkahloun AG, Lotufo RA, Choudary A, Dougherty ER, Suh E and Kallioniemi O. (2003). *Genome Res.*, **13,** 2341–2347.
Paddison PJ and Hannon GJ. (2002). *Cancer Cell*, **2,** 17–23 . RNA interference: the new somatic cell genetics?.

Paddison PJ and Hannon GJ. (2003). *Curr. Opin. Mol. Ther.*, **5,** 217–224.

Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, Scobie K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ and Hannon GJ. (2004). *Nature*, **428,** 427–431.

Shoemaker DD, Lashkari DA, Morris D, Mittmann M and Davis RW. (1996). *Nat. Genet.*, **14,** 450–456.

Shuman HA and Silhavy TJ. (2003). *Nat. Rev. Genet.*, **4,** 419–431.

Silva JM, Mizuno H, Brady A, Lucito R and Hannon GJ. (2004). *Proc. Natl. Acad. Sci.*, **101,** 6548–6552.

Somma MP, Fasulo B, Cenci G, Cundari E and Gatti M. (2002). *Mol. Biol. Cell*, **13,** 2448–2460.

St Johnston D. (2002). *Nat. Rev. Genet.*, **3,** 176–188.

Tabara H, Grishok A and Mello CC. (1998). *Science*, **282,** 430–431.

Timmons L and Fire A. (1998). *Nature*, **395,** 854.

Ziauddin J and Sabatini DM. (2001). *Nature*, **411,** 107–110.

A Genetic Screen for Candidate Tumor Suppressors Identifies REST

Thomas F. Westbrook[1], Eric S. Martin[2*], Michael R. Schlabach[1*], Yumei Leng[1], Anthony C. Liang[1], Bin Feng[2], Jean J. Zhao[3], Thomas Roberts[3], Gail Mandel[5], Gregory Hannon[6], Ronald DePinho[2], Lynda Chin[2,4], and Stephen J. Elledge[1,#]

[1]Howard Hughes Medical Institute, Department of Genetics, Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115

[3]Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Pathology, Harvard Medical School, Boston, MA 02115

[4]Department of Dermatology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115

[5]Howard Hughes Medical Institute, Department of Neurobiology and Behavior, State University of New York, Stony Brook, NY 11794

[6]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724

*These authors contributed equally to this work.

Running Title: Tumor Suppressor Screen

To whom correspondence should be addressed

Dr. Stephen J. Elledge
Department of Genetics
Partners Center for Genomics and Genetics
Harvard Medical School
77 Avenue Louis Pasteur
Boston, MA 02115, USA

Phone 617-525-4510
FAX 617-525-4500
e-mail: selledge@genetics.med.harvard.edu

Summary

Tumorigenesis is a multistep process characterized by a myriad of genetic and epigenetic alterations. Identifying the causal perturbations that confer malignant transformation is a central goal in cancer biology. Here we report an RNAi-based genetic screen for genes that suppress transformation of human mammary epithelial cells. We identified genes previously implicated in proliferative control and epithelial cell function including two established tumor suppressors, *TGFBR2* and *PTEN*. In addition, we uncovered a previously unrecognized tumor suppressor role for REST/NRSF, a transcriptional repressor of neuronal gene expression. Array-CGH analysis identified *REST* as a frequent target of deletion in colorectal cancer. Furthermore, we detect a frameshift mutation of the *REST* gene in colorectal cancer cells that encodes a dominantly acting truncation capable of transforming epithelial cells. Cells lacking REST exhibit increased PI(3)K signaling and are dependent upon this pathway for their transformed phenotype. These results implicate REST as a human tumor suppressor and provide a novel approach to identifying candidate genes that suppress the development of human cancer.

Introduction

The evolution of human cells into malignant derivatives is driven by the aberrant function of genes that positively and negatively regulate various aspects of the cancer phenotype, including altered responses to mitogenic and cytostatic signals, resistance to programmed cell death, immortalization, neoangiogenesis, and invasion and metastasis (Hanahan and Weinberg, 2000). The integrity of these gene functions is compromised by substantial genetic and epigenetic alterations observed in most cancer cell genomes. To understand the tumorigenic process, it is imperative to identify and characterize the genes that provide tumor cells with the capabilities requisite for their initiation and progression. However, the identities of those genes that contribute to the tumor phenotype are often concealed by the frequent alterations in genes that play no role in tumorigenesis.

Identifying genes that restrain tumorigenesis (tumor suppressors) has proven especially challenging due to their recessive nature. Further complicating their discovery are the multifaceted mechanisms by which tumor suppressor genes are inactivated including changes in copy number and structure, point mutations, and epigenetic alterations (Balmain et al., 2003). Moreover, the mechanisms by which tumor suppressor genes are inhibited may vary between tumors. With this in mind, a variety of molecular and cytogenetic technologies have been used to establish extensive catalogs of genetic alterations within human cancers (Albertson et al., 2003; Futreal et al., 2004). And while it is generally accepted that highly recurrent aberrations signify changes that are important for tumor development, the causal perturbations underlying tumor genesis are often confounded by the extensive size of alterations and the large number that are

incidental to the tumor phenotypes. As such, new strategies to delineate genes with functional relevance to tumor initiation and development are essential to understanding these processes.

One approach to this problem involves the use of *in vitro* models of human cell transformation. In such models, primary cells are transformed into tumorigenic derivatives by the coexpression of cooperating oncogenes (Elenbaas et al., 2001; Hahn et al., 1999; Zhao et al., 2004). These experimental models have been useful in delineating the minimum genetic perturbations required for transformation of various human cell types as well as evaluating the functional cooperation between a gene of interest and a defined genetic context. To date, these models of human cell transformation have incorporated genes already implicated in human tumorigenesis. However, such models also provide a potentially useful platform for the identification of new pathways that contribute to the transformed phenotype.

The emergence of RNA interference (RNAi) as a mechanism to silence gene expression has enabled loss-of-function analysis in mammalian cells in a potentially genome-wide manner (Berns et al., 2004; Paddison et al., 2004). We have utilized such an RNAi-based, forward genetic approach to identify genes that suppress oncogenic transformation in a defined human mammary epithelial cell model. We identified approximately 25 potential suppressors of epithelial cell transformation (SECT) genes that represent candidate tumor suppressors. Several are associated with known cancer-relevant pathways including Ras, PI(3)K and TGF-β signaling. In addition, we provide evidence that one of these candidates, the transcriptional repressor REST/NRSF, plays a

previously unappreciated role in tumor suppression. These findings support the utility of this novel approach to the identification of cancer relevant genes.

Results

RNAi Screen for Suppressors of Epithelial Cell Transformation

Recently, several cell culture models of human cell transformation have been described in which primary human cells are engineered to express combinations of dominantly-acting cellular and viral oncogenes and subsequently measured for anchorage-independent proliferation, an *in vitro* hallmark of transformation (Elenbaas et al., 2001; Hahn et al., 1999; Zhao et al., 2004). We sought to identify short-hairpin RNAs (shRNAs) that cooperate within the context of such a model. As >80% of cancers arise from epithelial tissues (AmericanCancerSociety, 2000), we chose to examine cells derived from human mammary epithelial cells (HMECs) to increase the probability of finding genes with relevance to epithelial cancers. The cells utilized for this screen (referred to as TLM-HMECs) were created by sequentially introducing the human telomerase catalytic subunit (hTERT) and the Large T antigen (LT) of SV40 into HMECs (Zhao et al., 2003). In addition, these cells exhibit elevated expression of the endogenous c-*myc* gene resulting from extended culture *in vitro* (Wang et al., 2000). Important to the efficacy of this screen, these cells do not proliferate in the absence of extracellular matrix (<1 colony in 10^4; Figure 1B, left panel) (Zhao et al., 2003). Recent experiments have demonstrated that hyper-activation of the PI(3)K pathway by ectopic expression of a PI(3)K mutant endows these cells with the ability to proliferate in an anchorage-

independent manner (Zhao et al., 2003), suggesting that TLM-HMECs are susceptible to transformation by a single genetic event. However, since recent evidence suggests that over-expressed oncogenes can confer different biological effects than oncogenes expressed at endogenous levels (Guerra et al., 2003; Tuveson et al., 2004), we sought to determine whether disruption of endogenous PI(3)K regulation is sufficient to elicit transformation. To this end, an shRNA directed against the PTEN tumor suppressor was introduced into TLM-HMECs by retroviral-mediated gene transfer. PTEN catalyzes the removal of the 3-position phosphate from PtdIns(3,4,5)P_3 and is a well-characterized antagonist of PI(3)K-dependent signals (Vivanco and Sawyers, 2002). PTEN-specific shRNA significantly reduced PTEN expression (Figure 1A). Importantly, reduced PTEN expression conferred robust anchorage-independent proliferation to a level similar to an activated mutant of PI(3)K (myr-p110α) (Figure 1B, right panel), thereby validating that RNAi-mediated loss-of-function of this tumor suppressor is capable of transforming TLM-HMECs.

To identify endogenous suppressors of epithelial cell transformation (SECT) genes (Figure 1C), we infected TLM-HMECs with a retroviral shRNA library we previously constructed in pSM1 (Paddison et al., 2004). This library consists of 28,000 sequence-verified shRNAs targeting ~9,000 genes, with each shRNA linked to a unique 60-nucleotide sequence (DNA "barcode"). These molecular barcodes can be used to monitor relative frequencies of individual shRNAs in complex populations via microarray technology (Hensel et al., 1995). TLM-HMECs infected with a control retrovirus or pSM1-library were assessed for anchorage-independent proliferation (Figure 1C). Only cells infected with the pSM1-library exhibited formation of macroscopic

colonies in semisolid media (Figure 1D, top panels), indicating the presence of shRNAs that transform TLM-HMECs. Approximately 100 anchorage-independent clones were pooled and analyzed for the enrichment of barcodes linked to the individual shRNAs (Figure 1D, bottom panel). To support the results from the hybridization studies, we sequenced approximately 200 individual anchorage-independent clones isolated from pSM1-transductants (including 70 colonies used for the barcode microarray hybridization). Sequencing of the proviral shRNAs from these colonies identified twenty-five unique shRNAs. Importantly, these approaches yielded similar results, with 18 of the 25 genes revealed by sequencing of individual clones also identified by barcode microarray analysis (Table S1). This is the first application of the barcode approach in a mammalian screen, and the high correspondence of identities with the sequenced clones indicates that such an approach harbors promise in more complex experimental designs.

Most of the shRNAs identified in this screen target genes known or predicted to function in signal transduction or transcriptional regulation (Table 1), consistent with the role of these gene classes in regulating complex cell behaviors. While the majority of these genes have not been directly examined for their relationship to cancer pathogenesis, several are implicated in the regulation of cancer-relevant pathways. Notably, we identified an shRNA targeting CAPRI (*RASA4*), a calcium-sensing Ras GTPase Activating Protein (Ras-GAP) previously shown to inhibit Ras-dependent signaling (Lockyer et al., 2001). This is consistent with the ability of a constitutively-active Ras mutant to transform TLM-HMECs (J.Z. and T.R., unpublished observations). δ-catenin (*CTNND2*) and K-cadherin (*CDH6*) were also isolated (Table 1), suggesting a potential role for adherens junctions in preventing epithelial cell transformation. Strikingly, the

type-II receptor for Transforming Growth Factor-β (TGF-β) was also identified, thus implicating the TGF-β tumor suppressor pathway in the control of epithelial cell transformation (see below).

The vast majority (90%) of colonies analyzed in this screen represented shRNAs directed against eight genes (unshaded in Table 1). Therefore, we focused our subsequent investigations on the gene targets of these frequently isolated and potentially more penetrant shRNAs. In order to validate putative gene targets, we assessed the transforming potential of shRNAs targeting different regions of each candidate. Parental TLM-HMECs were transduced with shRNA-encoding retroviruses and measured for anchorage-independent proliferation. Importantly, for five of six candidate genes tested, two shRNAs transformed TLM cells (Figure 1E, Figure 2B, and data not shown), indicating that the observed phenotype was most likely attributable to reduced function of the intended target and not caused by off-target effects. In contrast, two independent shRNAs directed against the pseudogene VDAC2P did not recapitulate transforming potential of the library-derived VDAC2P-shRNA (data not shown), implying that these shRNAs did not alter VDAC2P expression or that the library-derived shRNA targeted the expression of other genes underlying the transformed phenotype. Consistent with the latter hypothesis, VDAC2P expression was undetectable in TLM-HMECs (data not shown).

Endogenous TGF-β Signaling Suppresses Cellular Transformation

As negative regulators of oncogenic transformation, genes isolated in this screen may represent candidate tumor suppressor genes or impinge upon pathways critical to the

genesis of cancer. Notably, this approach identified the Transforming Growth Factor-β Receptor II (TGF-βRII) as a regulator of transformation (Table 1). Observations in mouse models and in human tumors indicate that the TGF-βRII (as well as several components of TGF-β signaling) is a potent tumor suppressor in numerous tissues including the mammary and colonic epithelia (Derynck et al., 2001; Siegel and Massague, 2003). The TGF-β pathway is a potent inhibitor of epithelial cell proliferation, but has not been previously implicated in regulating oncogenic transformation *in vitro*. To verify the role of TGF-βRII in suppression of transformation, two retroviral-encoded shRNAs targeting independent sequences within TGF-βRII were introduced into TLM-HMECs. These shRNAs reduced TGF-βRII expression levels and impaired phosphorylation of SMAD2, a substrate and transducer of endogenous TGF-β receptor signaling (Figure 2A). Importantly, TGF-βRII-targeted shRNAs also elicited robust anchorage-independent proliferation in TLM cells (Figure 2B), thus validating the identification of TGF-βRII in our screen. Reduced TGF-β signaling did not alter proliferation on an adhesive cell culture surface (Figure 2C), suggesting that loss of extracellular matrix (ECM) interactions may induce TGF-β signaling or alter the threshold of endogenous TGF-β signaling necessary to elicit a cytostatic response.

To further examine the role of endogenous TGF-β signaling in restraining cell transformation, we inhibited TGF-β signal transduction by alternative mechanisms and assessed the consequences on anchorage-independent proliferation. TLM-HMECs were transduced with retroviruses encoding a previously characterized dominant-negative mutant of TGF-βRII or SMAD7, a negative regulator of TGF-β receptor signaling

(Siegel and Massague, 2003). Expression of either cDNA conferred growth in semisolid media (Figure 2D), indicating that the transforming capacity of TGF-βRII shRNAs is not an RNAi-specific phenomenon. Conversely, ectopic activation of TGF-β signaling by a constitutively active mutant of TGF-βRI (T204D) was able to restrain anchorage-independent proliferation elicited by PTEN knockdown (Figure 2E), suggesting a genetic interaction between PI(3)K and TGF-β signaling in the context of cell transformation. This observation is interesting in light of evidence from several systems demonstrating that TGF-β and PI(3)K signals are integrated at multiple levels to regulate survival and proliferation (Conery et al., 2004; Remy et al., 2004; Seoane et al., 2004). Further investigation is required to determine the functional nodes through which these two pathways interact during HMEC transformation.

Inactivation of REST in Human Tumors

Consistent with a role in suppressing oncogenic transformation, four of the SECT genes we identified in the screen or initial experiments are either established tumor suppressors (TGF-βRII and PTEN) or regulators of cancer-relevant signaling machinery including the Ras proto-oncogene (*RASA4*) and cadherin complexes (*CTNND2*). Therefore, it is probable that other genes identified in our screen also represent candidate tumor suppressors and may be found altered in tumors. A hallmark of tumor suppressor loci is their high frequency of loss of heterozygosity (LOH) in tumors. Consequently, we examined whether our candidate genes reside in genetic loci targeted for such chromosomal aberrations in human tumors. Array-based comparative genomic hybridization (aCGH) has evolved into a high-throughput method for cataloguing such

copy number aberrations (CNAs) with high resolution. To this end, aCGH has been successfully utilized to characterize genomic alterations in the context of pancreatic adenocarcinoma(Aguirre et al., 2004), and more recently used to define CNAs in a large collection of human colon tumors and cell lines (Martin/DePinho, manuscript in preparation). As described(Aguirre et al., 2004), overlapping CNAs from individual colon tumor samples and tumor-derived cell lines were used to define minimal common regions (MCRs) of loss or gain. These discrete MCRs were further prioritized based on parameters of confidence and significance including: (1) recurrence in multiple independent samples, (2) high Log_2 ratio of change (e.g. depth of deletion), (3) focal nature (e.g. MCR $\leq$ 2.0Mb), and (4) MCR encompassing no more than 5 annotated genes (Aguirre et al., 2004). From this analysis, nine high-confidence MCRs of recurrent focal deletions were identified within the colon cancer genome, in sum representing only 34 annotated genes (Figure 3A). Consistent with its role in the pathogenesis of human cancers (Ruas and Peters, 1998), the $p16^{INK4A}$ tumor suppressor was present within one of these focal deletions (Figure 3A).

Remarkably, without *a priori* knowledge of the candidate genes listed in Table 1, this genomic approach identified high-confidence MCRs that included two highly penetrant candidates from our RNAi screen: *TGFBR2* and *REST* (Figure 3A). RE1-Silencing Transcription Factor (REST)/ Neuron Restrictive Silencing Factor (NRSF) is a transcriptional regulator best characterized for its role in repressing neuronal genes including neurotrophins and cell-adhesion molecules in non-neuronal tissues (Chong et al., 1995; Schoenherr and Anderson, 1995). Intriguingly, a variety of human tumors including those arising in breast, ovary, and lung activate expression of neuron-specific

genes. In some instances, the inappropriate expression of these neural genes elicits an autoimmune response that culminates in neurological disorders, collectively known as paraneoplastic neurologic degenerations (PND; see discussion) (Albert and Darnell, 2004). Such aberrant neural gene expression suggests that these cancers harbor defects in regulators of neuronal programs. Consequently, we investigated a potential role for REST in human tumor suppression. Deletions of varying size encompassing the *REST* locus on chromosome 4 were detected in a significant proportion of tumors, with evidence of genetic loss in 14 of 42 primary tumors and in 13 of 38 cell lines (Figure 3C), suggesting that chromosomal deletions targeting the *REST* locus are a frequent event in colon cancer. Importantly, microdeletions encompassing the *REST* gene were detected in a primary tumor specimen (CRC_19T) and a colon cancer cell line (LS123), thus defining a minimal common region. This MCR (Figure 3B, top panels) encompasses only five known genes, with the *REST* gene residing at the peak amplitude of each of these focal deletions (Figure 3B, bottom panels). Collectively, the unbiased identification of *REST* as a target of recurrent microdeletions as well as frequent larger deletions strongly suggest that *REST* is targeted for inactivation during colon cancer pathogenesis.

REST is widely expressed throughout non-neuronal tissues including the colonic epithelium (Supplemental Figure 1) (Chong et al., 1995; Schoenherr and Anderson, 1995). The above results suggest that loss of REST expression may confer a selective advantage during the evolution of tumor cells. This hypothesis predicts that cells with defective REST function may be sensitive to reconstitution of REST. We examined this prediction by ectopically expressing REST in colon cancer cells that have lost (SW1417) or retain (SW620) endogenous REST expression (Figure 4A). Exogenous REST

expression elicited a mild decrease in the proliferation of SW620 colon cancer cells (Figure 4B). In contrast, ectopic REST expression significantly reduced colony formation in SW1417 cells (>50-fold; Figure 4B), indicating that these cells are highly dependent on the absence of REST for their proliferation *in vitro*. Coupled with the function of REST in suppressing epithelial cell transformation (Figure 1E), these data strongly support a role for REST in tumor suppression.

In order to establish a more causal relationship between disruption of REST function and tumorigenesis, we analyzed primary colon tumors and colon tumor cell lines for the presence of mutations within the *REST* coding region. We sequenced exons 2-4 of the *REST* gene in a total of 86 colon cancers (48 tumors and 38 cell lines). We identified a single-nucleotide deletion within *REST* exon 4 of cells derived from a colorectal adenocarcinoma (DLD-1)(Figure 4A). This frameshift mutation results in a premature termination in the center of the coding sequence, yielding a protein with a predicted size of approximately 85kDa. This frameshift mutant, referred to as REST-FS, was detected by N-terminal REST antibodies in lysates from DLD-1 cells but not that of other colon cancer cell lines (Figure 4B, top panel). To confirm this, the frameshift mutation was engineered into a Flag-tagged REST cDNA and stably introduced into cells. Expression of an 85kDa protein was detected using REST- (Figure 4B, bottom panel) and FLAG-specific antibodies (Figure 4C, top panel).

Tumor cells harboring this frameshift mutation also encoded and expressed a wild-type REST allele (data not shown). Unfortunately, the primary tumor from which these cells were derived is not available, preventing analysis to determine whether the frameshift mutation occurred *in vivo*. However, we hypothesized if expression of REST-

FS was important to the genesis of this cancer, the mutant allele should exhibit dominant-negative activity. REST-FS lacks the C-terminal repressor domain that interacts with CoREST, a corepressor molecule that mediates transcriptional repression and silencing by REST(Andres et al., 1999). To determine if REST-FS displays properties of a dominant negative protein, we transduced the HMEC derivative (TLM cells) used in our initial screen with wild-type or mutant REST and assessed anchorage-independent proliferation. Cells expressing REST-FS, but not wild-type REST, exhibited robust colony formation (Figure 4C, bottom panel). The ability of REST-FS to phenocopy the transforming activity of REST shRNAs strongly suggests that this mutant interferes with the functions of endogenous REST that restrain epithelial cell transformation. Taken together, these data provide strong support for the hypothesis that REST is a tumor suppressor in human cells.

REST Suppresses PI(3)K Signaling

The implication that REST is involved in regulating the transformed state of epithelial cells led us to determine which molecular circuits might be affected by loss of REST function. Activation of PI(3)K-dependent signaling by a variety of mechanisms has been shown to confer transformation in HMECs (Zhao et al., 2003), indicating that this pathway provides a potent stimulus to the transformed phenotype. Furthermore, deregulation of PI(3)K signaling occurs in a wide spectrum of human cancers (Vivanco and Sawyers, 2002). As such, we examined the impact of disrupting REST function (Figure 5A) on the activation of the PI(3)K pathway. TLM-HMECs expressing control- or REST-shRNA were deprived of growth factors, restimulated with EGF for various

times, and subsequently analyzed for activating phosphorylation of Akt (serine-473), an essential effector of PI(3)K signaling. Stimulation of Akt phosphorylation by EGF was enhanced in cells expressing REST-shRNA throughout the timecourse (Figure 5B). Consistent with increased signaling downstream of Akt, phosphorylation of ribosomal S6 protein and translational inhibitor 4E-BP1 was also upregulated in cells expressing REST shRNA (Figure 5C). In agreement, recent experiments have demonstrated that ectopic expression of eIF-4E (the target of 4E-BP1) confers transformation to mammary epithelial cells and promotes tumor formation *in vivo* (Avdulov et al., 2004; Wendel et al., 2004). These results indicate that impaired REST function confers an increase in both the intensity and duration of PI(3)K-dependent signaling. In order to investigate whether PI(3)K signaling is required for REST-shRNA induced transformation, we utilized a dominant-negative mutant of the PI(3)K regulatory subunit, p85. This mutant (referred to as Δp85) has previously been shown to abolish H-Ras^{V12}-induced and SV40 st-induced cell transformation (Rodriguez-Viciana et al., 1997; Zhao et al., 2003). TLM-HMECs expressing control-, PTEN-, or REST-shRNAs were transduced with a control or Δp85-encoding retrovirus. Expression of Δp85 (Figure 5D, top panel) did not alter growth of cells on an adhesive cell culture surface (data not shown). Consistent with the role of PTEN as an antagonist of PI(3)K signaling, Δp85 abrogated anchorage-independent growth of PTEN-shRNA cells (Figure 5D, bottom panel). Similarly, REST-shRNA induced transformation was inhibited in the presence of Δp85 (Figure 5D, middle and bottom panels), indicating that PI(3)K signaling is required for transformation conferred by loss of REST. Coupled with the aberrant activation of PI(3)K-dependent signals in response to reduced REST expression, these results suggest that inhibition of

PI(3)K signaling might be a mechanism through which REST suppresses transformation and tumorigenesis.

Discussion

Suppressors of Epithelial Cell Transformation

In this study, we applied an shRNA-based genetic approach to identify genes suppressing oncogenic transformation of human epithelial cells. Within the context of this model, we have identified two previously established tumor suppressor genes (*PTEN* and *TGFBR2*) that actively restrain HMEC transformation. In addition, we have identified several genes that impinge upon pathways implicated in cancer pathogenesis. For example, we isolated the calcium-sensing Ras-GAP, CAPRI (*RASA4*), a previously described negative regulator of the Ras proto-oncogene (Lockyer et al., 2001). Consistent with a role for regulation of Ras signaling in our experimental system, activated Ras can transform TLM-HMECs. Likewise, identification of δ-catenin (*CTNND2*), a member of the p120 catenin family that regulates cadherin stability and trafficking (Reynolds and Roczniak-Ferguson, 2004), implicates a role for adherens junctions in constraining transformation of HMEC. Notably, disruption of adherens junction components has been shown to alter several growth-regulatory pathways (eg. β-catenin, PI(3)K) and has been linked to cancer progression in a variety of tissues (Cavallaro and Christofori, 2004; Vasioukhin et al., 2001).

The novelty of this genetic approach is in the unbiased identification of new and unanticipated tumor suppressor functions. In this regard, our studies provide significant evidence for the identification of a novel tumor suppressor, the transcriptional repressor REST/NRSF (see below).

Involvement of REST/NRSF in Human Cancer

Transcription factors often coordinately control complex programs of gene expression during development, and as such, are logical candidates underlying the aberrant activation of developmental programs in cancer. Here, we present several lines of evidence that REST may play a role in tumor suppression in humans. First, reduced REST function (mediated by RNAi or expression of dominant-negative REST) promotes transformation of human epithelial cells. Conversely, reconstitution of REST expression elicits a dramatic proliferation defect in colon cancer cells that have lost endogenous REST function. Strikingly, an independent aCGH-based search for genomic loci characterized by recurrent microdeletions identified the REST locus as a high-confidence target in colorectal cancer. This high-confidence list includes two previously established tumor suppressors, $p16^{INK4A}$ and TGFβRII (Derynck et al., 2001; Ruas and Peters, 1998; Siegel and Massague, 2003), the latter of which was also identified as a SECT gene. In addition, larger deletions encompassing the REST gene were frequent in colon tumors and tumor cell lines. Furthermore, we isolated a frameshift mutation of the REST coding region in colorectal tumor cells that displays properties of a dominant-negative, transforming human epithelial cells to a level similar to REST shRNAs. Finally, we show that impaired REST function enhances the intensity and duration of PI(3)K signaling, a pathway that is aberrantly activated in many if not all human cancers(Vivanco and Sawyers, 2002). Importantly, PI(3)K activity was required for cellular transformation conferred by reduced REST function, indicating that suppression of PI(3)K signaling may be an important mechanism underlying the ability of REST to restrain the transformed state. The mechanism by which REST inhibits this oncogenic pathway is not yet clear. REST regulates a complex transcriptional program, and as such,

may impinge on PI(3)K signaling through multiple networks. However, it should be noted that BDNF and other neurotrophins are among the transcriptional targets repressed by REST. BDNF activates the TrkB receptor and has recently been shown to suppress anoikis via PI(3)K-dependent pathways (Douma et al., 2004), providing a plausible mechanism for activation of PI(3)K signaling in the absence of REST. Taken together, these data provide compelling support for a role for REST in human tumorigenesis and further validate the genetic approach we have undertaken.

Deregulation of neuronal programs has been implicated in cancer through the phenomena of paraneoplastic neurological degenerations (PNDs) (Albert and Darnell, 2004). In these diverse neurological disorders, tumors originating in non-neuronal tissues activate expression of neural peptides that elicit an immune response to the tumor as well as the host nervous system. Although the expression of some of these neural antigens is common or universal in some cancers (breast, ovarian, small-cell lung cancers) (Albert and Darnell, 2004), the occurrence of PNDs in cancer patients is rare, suggesting that other factors play a role in the autoimmune response. Nonetheless, aberrant activation of neuronal gene expression raises the possibility that regulators of neurogenesis are malfunctioning in these tumors. To investigate this possibility, we searched for previously identified PND antigens among known REST targets and discovered RE1-binding sites in several PND antigen encoding genes. Furthermore, promoters of two of these PND antigens (synaptotagmin and glutamic acid decarboxylase) were recently shown to be directly bound by REST, and moreover, transcription of synaptotagmin was induced upon expression of a dominant-negative REST mutant similar to the one we identified in tumors (Ballas and Mandel, in press). This supports the hypothesis that

defects in the REST pathway are tied to PND pathology and tumorigenesis. In further support of a role in tumorigenesis, REST expression was recently shown to be absent in a subset of SCLC cells (Neumann et al., 2004). Additionally, microarray profiling also demonstrates down-regulation of REST expression in prostate and small cell lung cancers, two malignancies that often display distinct neuroendocrine phenotypes (Dhanasekaran et al., 2001; Garber et al., 2001; Rhodes et al., 2004). These findings suggest that REST may play a broader role in human malignancies in addition to colorectal cancer.

Genetic Screening for Tumor Suppressor Pathways

While further studies will be needed to determine the extent to which REST, or other genes we have identified, are involved in human cancer, our results point to the utility of this approach for identifying recessive cancer relevant genes. Although many of the genes isolated via this strategy may not be frequent targets of mutation in tumors, they may nonetheless reveal novel pathways relevant to tumorigenesis. While we have identified several suppressors of cellular transformation, it is clear that this screen was not saturated, and consequently, there remains significant potential for this method in the future discovery of SECT genes. Underscoring this, the shRNA library used in our screen was designed to target only ~9,000 genes, representing less than one-third of the annotated genes in the human genome. Furthermore, only one shRNA corresponding to any given SECT gene was isolated in this screen. For several of these genes, multiple corresponding hairpins were present within the pSM1 library. This suggests that many SECT candidates were not identified, because this library lacked a sufficiently penetrant

shRNA to elicit the transformed phenotype. This is not surprising since the pSM1 library was constructed before many of the parameters affecting siRNA efficiency had emerged. For instance, the stability of siRNA ends has been shown to bias the incorporation of sense/anti-sense strands into the RISC complex (Khvorova et al., 2003; Schwarz et al., 2003). Such siRNA design "rules" improve the potency of target gene suppression and will undoubtedly be incorporated into future generations of mammalian shRNA libraries, thus providing more potent tools in identifying SECT genes.

SECT candidates identified in this screen were selected for their transforming capacity in cooperation with the genetic milieu of TLM-HMECs (ectopic expression of hTERT, SV40 LT, and elevated endogenous c-myc expression). However, distinct classes of SECT genes are likely to be revealed in models incorporating alternative combinations of genetic perturbations. Additional SECT candidates may also be discovered in transformation models of alternative cell types, reflecting the different signaling requirements in cells derived from various human tissues (Hamad et al., 2002; Rangarajan et al., 2004), or by interrogating different cancer relevant phenotypes such as invasion, migration or angiogenesis. Indeed, as *in vitro* models of human cell transformation are engineered to more accurately reflect the molecular changes and heterotypic cellular complexity found in human cancers, we anticipate that this general strategy will enable a much more complete understanding of the pathways and processes that cancer cells usurp during tumorigenesis.

Experimental Procedures

Vectors and retroviral production

The shRNA library constructed in pSM1 has previously been described (Paddison et al., 2004). The PTEN shRNA-encoding pSuperRetro vector (accession NM_000314) and the Smad7-expression plasmid were kindly provided by Dennis McCance and Xiao-Jing Wang, respectively. For validation experiments, shRNAs targeting ZNF134, REST, CTNND2, VDAC2P, CDH6, and TGF-βRII were designed using the RNAi-design algorithm at http://katahdin.cshl.org:9331/portal/scripts/main2.pl. cDNAs encoding TGF-βRIIΔCYT and TGF-βRI(T204D) were generously provided by Joan Massague. The flag-tagged REST-FS cDNA was made by site-directed mutagenesis using QuickChange (Stratagene). TGF-βRIIΔCYT, TGF-βRI(T204D), REST, and REST-FS cDNAs were subcloned into the LPC retroviral vector. pWZL-neo-Δp85 has been previously described (Zhao et al., 2003). Retroviral supernatants were generated by transient transfection of Phoenix cells with the indicated retroviral and VSV-G-expression plasmids and collected 48 hours post-transfection.

Cell Culture

HMECs expressing hTERT and SV40 LT (TLM-HMECs) (Zhao et al., 2003) were cultured in mammary epithelial growth medium (MEGM, Cambrex). Colon cancer cell lines were maintained in DMEM or RPMI supplemented with 10% FBS and 50 μg/mL gentamycin. Stable cell lines were generated by transduction with the indicated retroviral supernatants in the presence of 8 μg/mL polybrene, with transduced cells selected for resistance to the appropriate drug: puromycin (2.0 μg/mL), neomycin (200 μg/mL).

Anchorage-independent proliferation assays were performed as described (Zhao et al., 2003) except cells were suspended in a top layer of 2.0% methylcellulose (Sigma) in MEGM. For each assay, the average of at least 3 experiments $\pm$ SD is shown. For growth curves, cells were seeded at a density of 5.0 X 10^4 per well in 6-well plates and cultured in MEGM. Cells were trypsinized and counted at the indicated time points (in triplicate with average $\pm$ SD shown). EGF restimulation experiments were performed as previously described (Zhao et al., 2003).

shRNA screen

TLM-HMECs were infected with a retroviral shRNA library constructed in pSM1 consisting of 28,000 shRNAs directed against ~9,000 human genes (Paddison et al., 2004) at an M.O.I. of 0.2. Transduced cells were selected for puromycin-resistance, seeded into semisolid media (as described above) at a density of 1.0 X 10^5 cells per 100mm plate (20 plates), and cultured for 3 weeks. Individual anchorage-independent clones were isolated and expanded on adhesive tissue culture dishes. A portion of the provirus (including shRNA cassette) from each clone was PCR amplified from genomic DNA using primers directed against sequences within the MSCV backbone and PGK promoter (forward: 5'-ctccctttatccagccctcac-3'; reverse: 5'- gagacgtgctacttccatttgtc-3'), and sequenced using an internal primer: 5'-gagggcctatttcccatgat-3'. EcoRI-XhoI fragments of the PCR-fragments were also subcloned into MCSV-SIN for validation experiments. In parallel, anchorage-independent clones were pooled, and genomic DNA was isolated using Qiagen DNeasy kit. PCR-amplification and transcription of barcodes,

hybridization of custom barcode microarrays (Agilent), and microarray analysis was performed as previously described (Paddison et al., 2004).

Immunoblotting

Cells were lysed in HLB buffer (50mM Tris, pH 8.0, 150mM NaCl, 1.0% Triton-X 100, 1mM DTT, supplemented with phosphatase- and protease-inhibitor cocktails (Calbiochem)) for 30 minutes with sonication. Western blotting was performed with the following antibodies: αPTEN (Upstate 07-016), αTGF-βRII (Upstate 06-227 and 06-318), αRan (BD Biosciences 610340), αDDB1 (AbCam Ab9194), αFlag (Sigma A8592), αp-465/467-Smad2 (Cell Signaling 3101), αSmad2 (Cell Signaling 3102), αp473-Akt (Cell Signaling 9271), αAkt (Cell Signaling, 9272), αp70-4EBP1 (Cell Signaling, 9455), α4EBP1 (Cell Signaling, 9452), αp235/236-S6 (Cell Signaling, 2211), αS6 (Cell Signaling, 2217), αRest-N and αRest-C (G. Mandel).

Array-CGH profiling and mutation analysis of REST

Genomic DNA extracted from 42 primary colon tumors and 38 colon tumor-derived cell lines using Puregene DNA extraction (Gentra Systems) was fragmented and random-prime labeled according to published protocols (Aguirre et al., 2004; Bachoo et al., 2004; Pollack et al., 1999) with modifications (For details, see http://genomic.dfci.harvard.edu/array_cgh.htm). Subsequently, fluorescence ratios of scanned images of the arrays were calculated. Raw array-CGH profiles were processed to identify statistically significant transitions in copy number using a segmentation algorithm which employs permutation analysis to determine the significance of change

points in the raw data (Aguirre et al., 2004) (Bachoo et al., 2004). Each segment is assigned a Log_2 ratio that is the median of the contained probes. Subsequently, data was imported into the dCHIP v1.3 software package for microarray analysis (http://www.dchip.org/) for visualization of the segmented data. For mutation analysis, coding exons of the *REST* gene were PCR amplified, purified, and sequenced using standard conditions (primers available upon request).

Acknowledgements

We are grateful to Joan Massague, Dennis McCance, Barry Thrash, and Xiao-Jing Wang for the gift of reagents. We recognize the Harvard Partners Genome Center at HPCGG for the sequencing of human tumor DNA. We thank Mamie Li, Zhenming Zhou, and Nancy Ryan for technical support and Frank Stegmeier, Guang Hu, Robert McDonald, Jennifer Hackett, and Don Nguyen for critical reading of the manuscript. We are also grateful to Joan Massague, Robert Darnell and Matthew Meyerson and members of the Elledge and Hannon laboratories for helpful discussions. T.F.W. is a fellow of the Susan G. Komen Foundation and is supported by grant PDF0403175. This work was supported by grants from NCI 5T32CA09361 (E.S.M.), NCI MMHCC U01 CA084313 (R.A.D.), RO1CA93947 and RO1CA99041 (L.C.), 5 P50 CA090381-03 (J.J.Z.), 5 P01 CA89021-04 and D.O.D.17-02-1-0692 (T.M.R.), grants from the NIH and DOD (G.J.H.), and a U.S. Army Innovator Award (W81XWH0410197) to S.J.E. R.A.D. is an American Cancer Society Research Professor. G.M. and S.J.E. are investigators with the Howard Hughes Medical Institute.

References

References

Aguirre, A. J., Brennan, C., Bailey, G., Sinha, R., Feng, B., Leo, C., Zhang, Y., Zhang, J., Gans, J. D., Bardeesy, N., *et al.* (2004). High-resolution characterization of the pancreatic adenocarcinoma genome. Proc Natl Acad Sci U S A *101*, 9067-9072.

Albert, M. L., and Darnell, R. B. (2004). Paraneoplastic neurological degenerations: keys to tumour immunity. Nat Rev Cancer *4*, 36-44.

Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003). Chromosome aberrations in solid tumors. Nat Genet *34*, 369-376.

AmericanCancerSociety (2000). Cancer Facts and Figures 2000 (Atlanta, American Cancer Society).

Andres, M. E., Burger, C., Peral-Rubio, M. J., Battaglioli, E., Anderson, M. E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. Proc Natl Acad Sci U S A *96*, 9873-9878.

Avdulov, S., Li, S., Michalek, V., Burrichter, D., Peterson, M., Perlman, D. M., Manivel, J. C., Sonenberg, N., Yee, D., Bitterman, P. B., and Polunovsky, V. A. (2004). Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant phenotype in human mammary epithelial cells. Cancer Cell *5*, 553-563.

Bachoo, R. M., Kim, R. S., Ligon, K. L., Maher, E. A., Brennan, C., Billings, N., Chan, S., Li, C., Rowitch, D. H., Wong, W. H., and DePinho, R. A. (2004). Molecular diversity of astrocytes with implications for neurological disorders. Proc Natl Acad Sci U S A *101*, 8384-8389.

Balmain, A., Gray, J., and Ponder, B. (2003). The genetics and genomics of cancer. Nat Genet *33 Suppl*, 238-244.

Berns, K., Hijmans, E. M., Mullenders, J., Brummelkamp, T. R., Velds, A., Heimerikx, M., Kerkhoven, R. M., Madiredjo, M., Nijkamp, W., Weigelt, B., *et al.* (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature *428*, 431-437.

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer *4*, 118-132.

Chong, J. A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J. J., Zheng, Y., Boutros, M. C., Altshuller, Y. M., Frohman, M. A., Kraner, S. D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell *80*, 949-957.

Conery, A. R., Cao, Y., Thompson, E. A., Townsend, C. M., Jr., Ko, T. C., and Luo, K. (2004). Akt interacts directly with Smad3 to regulate the sensitivity to TGF-beta induced apoptosis. Nat Cell Biol *6*, 366-372.

Derynck, R., Akhurst, R. J., and Balmain, A. (2001). TGF-beta signaling in tumor suppression and cancer progression. Nat Genet *29*, 117-129.

Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. (2001). Delineation of prognostic biomarkers in prostate cancer. Nature *412*, 822-826.

Douma, S., Van Laar, T., Zevenhoven, J., Meuwissen, R., Van Garderen, E., and Peeper, D. S. (2004). Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. Nature *430*, 1034-1039.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L., Popescu, N. C., Hahn, W. C., and Weinberg, R. A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev *15*, 50-65.

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M. R. (2004). A census of human cancer genes. Nat Rev Cancer *4*, 177-183.

Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. I., *et al.* (2001). Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A *98*, 13784-13789.

Guerra, C., Mijimolle, N., Dhawahir, A., Dubus, P., Barradas, M., Serrano, M., Campuzano, V., and Barbacid, M. (2003). Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. Cancer Cell *4*, 111-120.

Hahn, W. C., Counter, C. M., Lundberg, A. S., Beijersbergen, R. L., Brooks, M. W., and Weinberg, R. A. (1999). Creation of human tumour cells with defined genetic elements. Nature *400*, 464-468.

Hamad, N. M., Elconin, J. H., Karnoub, A. E., Bai, W., Rich, J. N., Abraham, R. T., Der, C. J., and Counter, C. M. (2002). Distinct requirements for Ras oncogenesis in human versus mouse cells. Genes Dev *16*, 2045-2057.

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., and Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. Science *269*, 400-403.

Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Lockyer, P. J., Kupzig, S., and Cullen, P. J. (2001). CAPRI regulates Ca(2+)-dependent inactivation of the Ras-MAPK pathway. Curr Biol *11*, 981-986.

Neumann, S. B., Seitz, R., Gorzella, A., Heister, A., Doeberitz, M. K., and Becker, C. M. (2004). Relaxation of glycine receptor and onconeural gene transcription control in NRSF deficient small cell lung cancer cell lines. Brain Res Mol Brain Res *120*, 173-181.

Paddison, P. J., Silva, J. M., Conklin, D. S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., *et al.* (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427-431.

Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet *23*, 41-46.

Rangarajan, A., Hong, S. J., Gifford, A., and Weinberg, R. A. (2004). Species- and cell type-specific requirements for cellular transformation. Cancer Cell *6*, 171-183.

Remy, I., Montmarquette, A., and Michnick, S. W. (2004). PKB/Akt modulates TGF-beta signalling through a direct interaction with Smad3. Nat Cell Biol *6*, 358-365.

Reynolds, A. B., and Roczniak-Ferguson, A. (2004). Emerging roles for p120-catenin in cell adhesion and cancer. Oncogene *23*, 7947-7956.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

Rodriguez-Viciana, P., Warne, P. H., Khwaja, A., Marte, B. M., Pappin, D., Das, P., Waterfield, M. D., Ridley, A., and Downward, J. (1997). Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. Cell *89*, 457-467.

Ruas, M., and Peters, G. (1998). The p16INK4a/CDKN2A tumor suppressor and its relatives. Biochim Biophys Acta *1378*, F115-177.

Schoenherr, C. J., and Anderson, D. J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science *267*, 1360-1363.

Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Seoane, J., Le, H. V., Shen, L., Anderson, S. A., and Massague, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell *117*, 211-223.

Siegel, P. M., and Massague, J. (2003). Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. Nat Rev Cancer *3*, 807-821.

Tuveson, D. A., Shaw, A. T., Willis, N. A., Silver, D. P., Jackson, E. L., Chang, S., Mercer, K. L., Grochow, R., Hock, H., Crowley, D., *et al.* (2004). Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell *5*, 375-387.

Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B., and Fuchs, E. (2001). Hyperproliferation and defects in epithelial polarity upon conditional ablation of alpha-catenin in skin. Cell *104*, 605-617.

Vivanco, I., and Sawyers, C. L. (2002). The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer *2*, 489-501.

Wang, J., Hannon, G. J., and Beach, D. H. (2000). Risky immortalization by telomerase. Nature *405*, 755-756.

Wendel, H. G., De Stanchina, E., Fridman, J. S., Malina, A., Ray, S., Kogan, S., Cordon-Cardo, C., Pelletier, J., and Lowe, S. W. (2004). Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. Nature *428*, 332-337.

Zhao, J. J., Gjoerup, O. V., Subramanian, R. R., Cheng, Y., Chen, W., Roberts, T. M., and Hahn, W. C. (2003). Human mammary epithelial cell transformation through the activation of phosphatidylinositol 3-kinase. Cancer Cell *3*, 483-495.

Zhao, J. J., Roberts, T. M., and Hahn, W. C. (2004). Functional genetics and experimental models of human cancer. Trends Mol Med *10*, 344-350.

Figure Legends

Figure 1. Identification of suppressors of epithelial cell transformation. (A) Cell lysates from TLM-HMECs expressing control- or PTEN-shRNAs were immunoblotted for expression of PTEN. The asterisk denotes a cross-reacting band that serves as a loading control. (B) Cells from (A) were cultured in semisolid media for 3 weeks and photographed at 20x (top panels) or 1x (bottom panels) magnification. Quantitation of macroscopic colony (>0.2mm) formation in semisolid media by TLM-HMECs transduced with empty retrovirus or virus encoding control shRNA, myr-p110α cDNA, or PTEN-shRNA. (C) Outline of the transformation screen. Details provided in Methods. (D) TLM-HMECs transduced with empty or pSM1-shRNA library and cultured in semisolid media for 3 weeks, photographed at 20x (top panels). Bottom panel illustrates a section of a barcode microarray. Barcodes were PCR-amplified from genomic DNA isolated from a pool of 100 anchorage-independent colonies. cRNA was transcribed from a total library preparation (red channel, 635nm) or from PCR-amplified barcode (green channel, 532nm) and hybridized to a barcode microarray. Enriched barcodes (green/yellow) are indicated by arrows. (E) TLM-HMECs expressing shRNAs targeting candidate genes (2 independent shRNAs per gene target) were cultured in semisolid media and quantitated for formation of macroscopic colonies. Experiments were performed in triplicate.

Table 1: Suppressors of epithelial cell transformation. This table lists gene targets of unique, sequence-verified shRNAs identified in 200 anchorage-independent colonies isolated from the screen. shRNAs identified in the context of double integrations (7 in

total) were disregarded. 90% of isolated anchorage-independent colonies encoded one of eight shRNAs (unshaded). For candidate validation, multiple shRNAs directed against independent sequences within a gene target were tested for transformation. (ND=not determined).

Figure 2. TGF-β signaling suppresses epithelial cell transformation. (A) Cell lysates from TLM-HMECs expressing control- or TGF-βRII-shRNAs were immunoblotted for expression of TGF-βRII, Ran (loading control), serine-465/467- phosphorylated SMAD2, or total SMAD2. (B) Cells from (A) were cultured in semisolid media for 3 weeks, photographed at 20x, and quantitated for formation of microscopic colonies. (C) Growth curves of cells expressing control (○), or two independent TGF-βRII-shRNAs (■ or ▲) grown in monolayer culture. (D) TLM-HMECs were stably transfected with empty plasmid (pcDNA3) or plasmid expressing SMAD7 or transduced with empty retrovirus (LPC) or retrovirus encoding a dominant-negative mutant of TGF-βRII (TGFβRII-ΔCYT). Resulting polyclonal cell lines were assayed for anchorage-independent proliferation. (E) TLM-HMECs were transduced with control (LPC) or PTEN-shRNA encoding retrovirus. PTEN-shRNA expressing cells were subsequently infected with empty retrovirus or virus expressing a constitutively active mutant of TGFβRI (TGFβRI-T204D). Each derivative cell line was assessed for proliferation in semisolid media.

Figure 3. Deletion of REST in human tumors. (A) High-confidence recurrent CNAs in colon cancer. Recurring focal CNAs in colon adenocarcinoma and tumor cell derivatives were identified and prioritized as previously described(Aguirre et al., 2004) with

modifications. Each of the identified minimum common regions (MCR) consists of <2.0Mb and ≤ 5 annotated genes. Peak probe values represent the minimum absolute $\log_2$-ratio detected within each MCR. Also shown are the chromosomal location, the number of NCBI annotated genes residing within the MCR, and candidate tumor suppressor genes within each MCR. SECT genes are highlighted in blue. (B) aCGH profiles of chromosome 4 in a primary tumor (19T) and a colon cancer cell line (LS123) showing deletion of a discrete region, with MCR defined by vertical red lines. Raw data and segmentation analysis are represented by blue circles and red horizontal lines, respectively. Peak data point for each deletion occurs at the REST probe (indicated by arrows). (C) Segmentation data of aCGH profiles from each of 42 primary colon tumors and 38 colon tumor-derived cell lines. Chromosomal gain or loss within a section of chromosome 4 is represented by a color gradient (red=gain, blue=loss). Gray boxes represent uninformative probe hybridizations.

Figure 4. Aberrant REST function in colon cancer. (A) REST expression was analyzed in cell lysates from SW620 and SW1417 colon cancer cell lines by immunoblotting with antibodies recognizing the C-terminus of REST. (B) SW620 and SW1417 cells were transduced with control- or REST-expressing retrovirus, seeded at 5,000 cells per dish, and cultured for 2 weeks under puromycin selection. (C) Single nucleotide deletion within exon 4 of the *REST* gene (top schematic) was observed in DLD-1 cells by sequence analysis (coding exons shown in gray). The premature termination (indicated by red x, bottom schematic) results in a protein retaining the DNA-binding domain (DBD) and N-terminal SIN3-binding domain, but lacking the repressor domain

responsible for interaction with CoRest (CBD). (D) The presence of truncated REST was analyzed by immunoblotting with antibodies recognizing the N-terminus of REST (top panel) or DDB1 (loading control, middle panel) against cell lystates from colon cancer cell lines. Protein lysates derived from HMECs transfected with empty vector (pcDNA3) or vector encoding the frameshift REST mutant (REST-FS) were probed with antibodies raised against the N-terminus of REST (bottom panel). (E) TLM-HMECs transduced with control retrovirus or retroviruses encoding flag-REST or flag-REST-FS were analyzed for exogenous cDNA expression using flag-specific antibodies (top panel) or assayed for anchorage-independent proliferation (bottom panel).

Figure 5. Loss of REST function stimulates PI(3)K signaling and requires it for HMEC transformation. (A) Cell lysates from TLM-HMECs expressing control- or REST-shRNAs were immunoblotted for expression of Akt (loading control, bottom panel) or REST (C-terminal antibody, top panel). (B) TLM-HMECs expressing control or REST-shRNA were starved in basal media without growth factors for 20 hours and restimulated with 10ng/mL EGF for indicated times. Cell lysates were immunoblotted with antibodies recognizing p-Akt (ser473) or total Akt. Akt phosphorylation was measured as a ratio of p-Akt/total Akt and represented as a percentage of phosphorylated Akt at 15 minutes of EGF stimulation in control-shRNA cells (control-shRNA □, REST-shRNA●). (C) Cell lysates from (A) were probed with antibodies recognizing p-S70 of 4E-BP1, total 4E-BP1, p-S235 and p-S236 of S6, and total S6 as indicated. (D) REST loss requires PI3K signaling for transformation. TLM-HMECs expressing control, REST, or PTEN-shRNA were transduced with empty or Δp85-expressing retrovirus. Immunoblot analysis of

endogenous p85 and exogenous Δp85 as indicated (top panel). Cells were analyzed for anchorage-independent proliferation (middle and bottom panels).

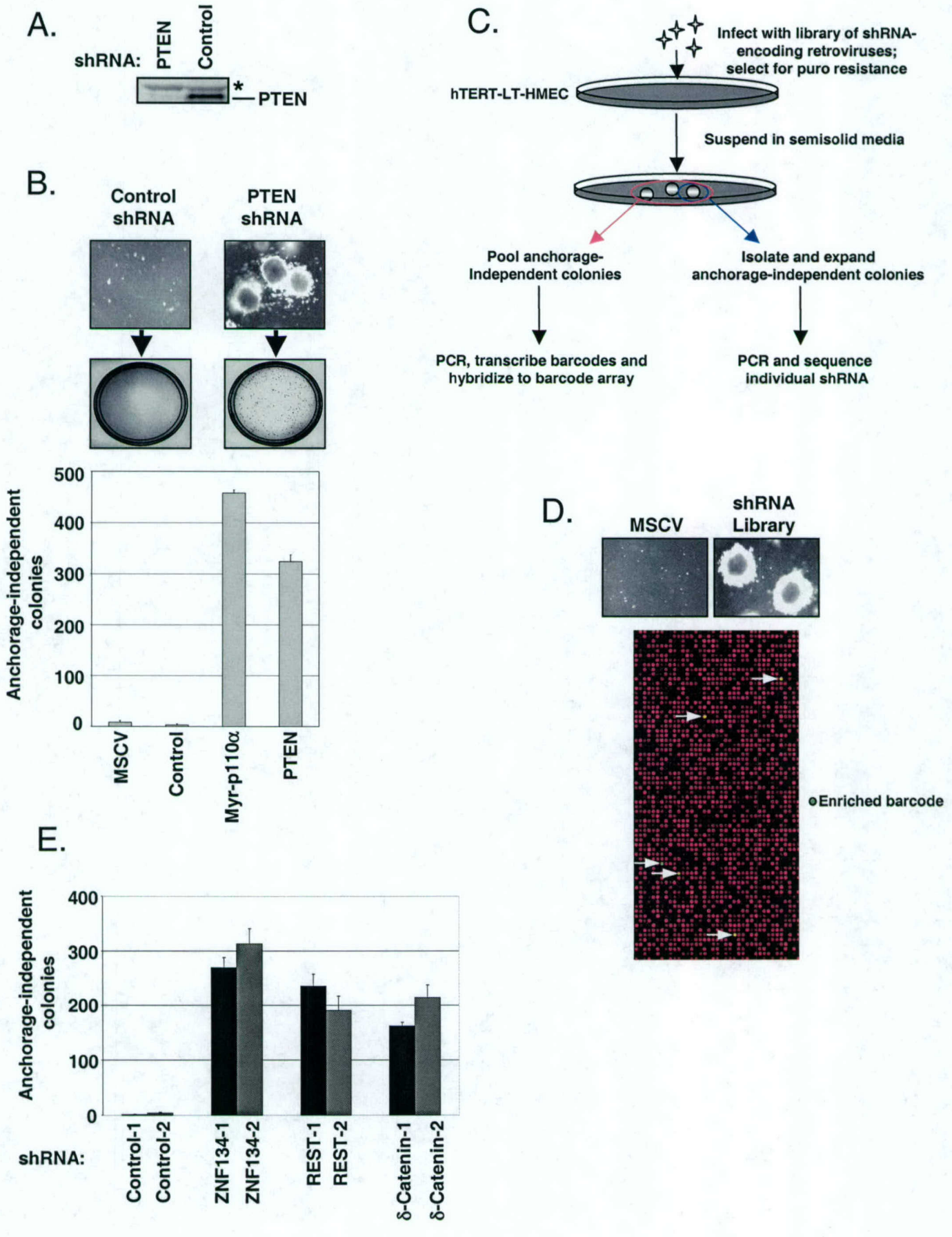

A.
shRNA:
PTEN
Control
*
PTEN
B.
Control shRNA
PTEN shRNA
Anchorage-independent colonies
500
400
300
200
100
0
MSCV
Control
Myr-p110α
PTEN
C.
Infect with library of shRNA-encoding retroviruses; select for puro resistance
hTERT-LT-HMEC
Suspend in semisolid media
Pool anchorage-Independent colonies
Isolate and expand anchorage-independent colonies
PCR, transcribe barcodes and hybridize to barcode array
PCR and sequence individual shRNA
D.
MSCV
shRNA Library
Enriched barcode
E.
Anchorage-independent colonies
400
300
200
100
0
shRNA:
Control-1
Control-2
ZNF134-1
ZNF134-2
REST-1
REST-2
δ-Catenin-1
δ-Catenin-2

Table 1

Gene	Previously Known Functions	Validated
CDH6	Type II cadherin; cell-cell adhesion and communication	+
CTNND2	Stabilization of adherens junctions	+
INPP4B	PIP2 phosphatase	ND
RASA4	Ras GAP; calcium-responsive inhibitor of Ras signaling	ND
REST	Transcriptional repression of neural genes	+
TGFBRII	TGF-β signaling; cytostatic and apoptotic programs in epithelial tissues	+
VDAC2P	None	-
ZNF134	None	+
BCL9	WNT/β-catenin signaling	ND
MAP4K4	TNFα-signaling; JNK activation	ND
PKN2	Rho signaling; Akt inhibition	ND
BDKRB2	G-protein coupled receptor	ND
LMO4	Transcriptional regulation; mammary gland development	ND
HAND1	Transcriptional regulation; cardiac morphogenesis	ND
AKT2	PI(3)K effector; survival signaling	ND
STAG3	Meiosis cohesin	ND
DUT	Nucleotide metabolism	ND
RPP30	tRNA processing	ND

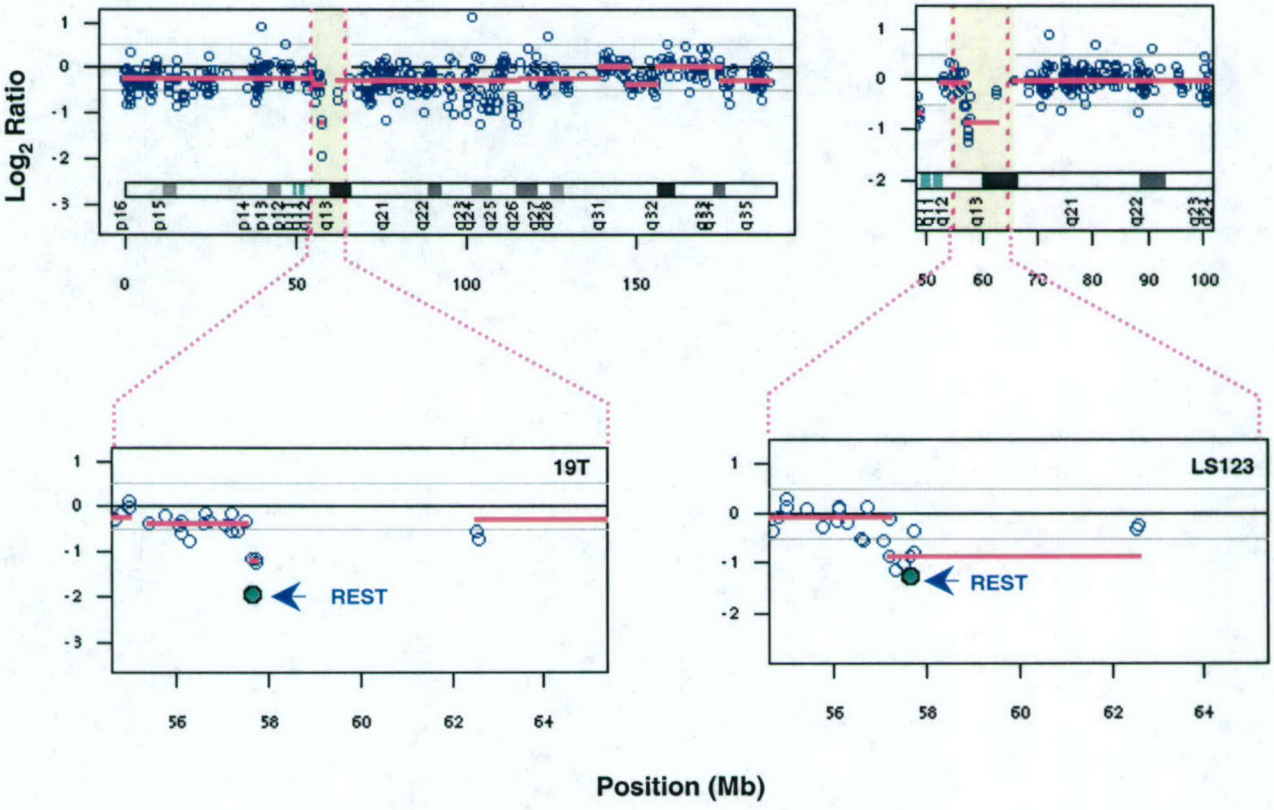
A
TGF-βRII
Ran
p-Smad2
Smad2
Control
TGF-βRII shRNA-1
TGF-βRII shRNA-2
B
Control
TGF-βRII shRNA-1
TGF-βRII shRNA-2
Anchorage-independent colonies
500
400
300
200
100
0
Control
TGF-bRII shRNA-1
TGF-bRII shRNA-2
C
Cell Number (x10^5)
10
5
0
1
2
3
4
Days
D
Anchorage-independent colonies
200
150
100
50
0
pcDNA
SMAD7
LPC
TGFβRII-ΔCYT
E
Anchorage-independent colonies
200
150
100
50
0
LPC
PTEN shRNA + vector
PTEN shRNA + TGFβRI-T204D

A

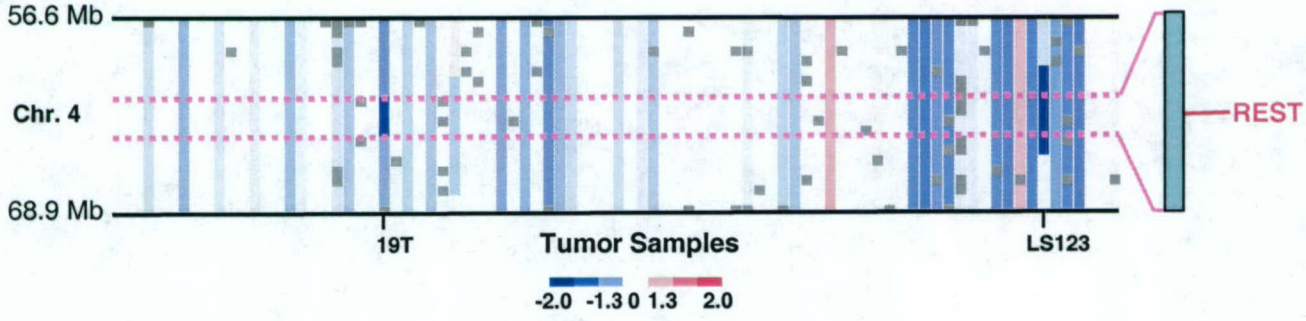
High confidence loci in colon cancer

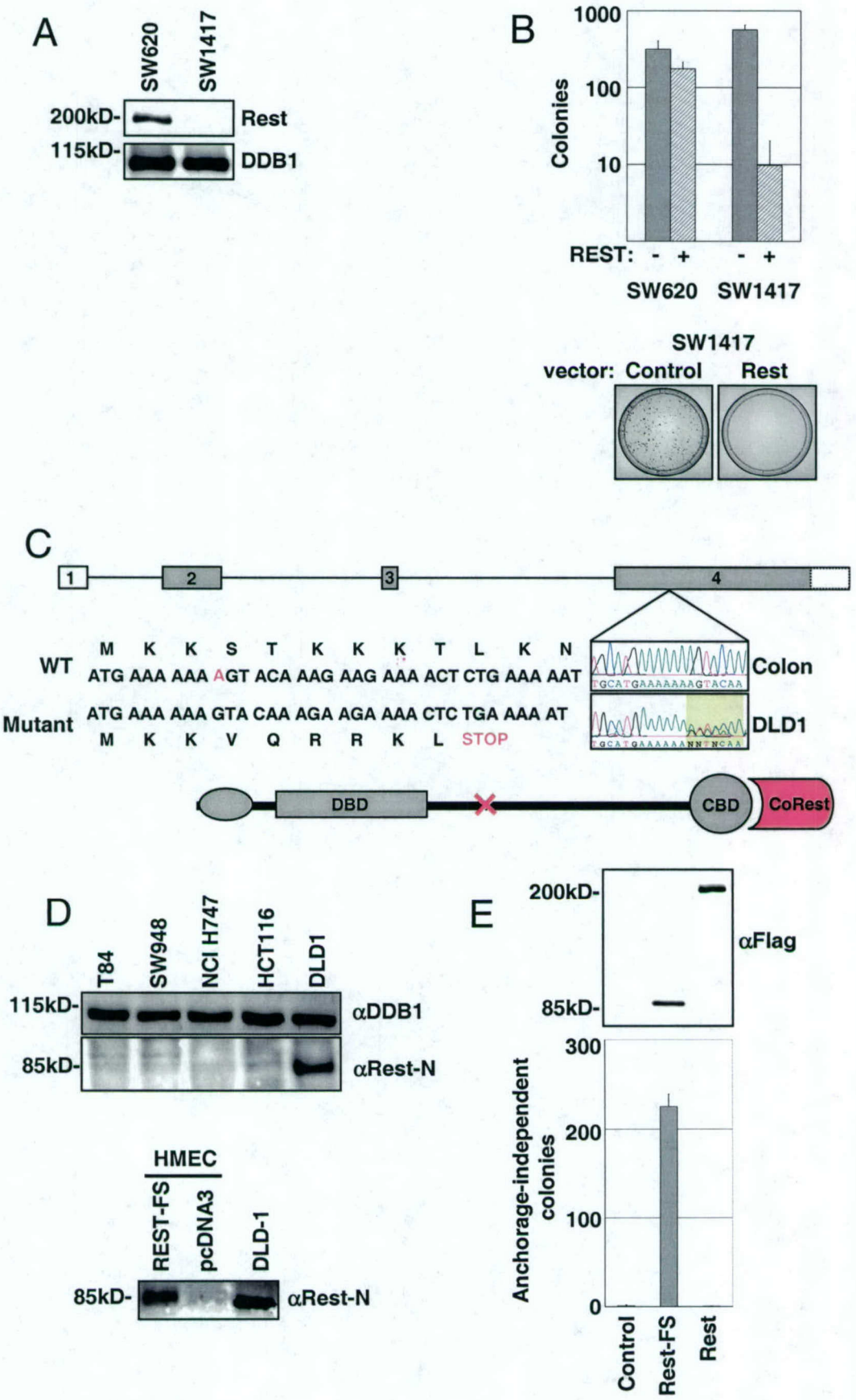
Chromosome	Minimum Common Region (Mb)	Peak Probe Value (log_2ratio)	# of Genes	Candidate Tumor Suppressor
3	1.66	-2.98	4	TGFBR2
4	0.22	-1.22	5	REST
8	0.41	-1.31	4	
9	0.55	-1.22	4	CDKN2A ($p16^{INK4A}$)
10	0.84	-1.24	2	
11	0.08	-4.12	2	
14	0.61	-1.57	5	
17	0.23	-3.4	5	
21	0.15	-1.36	3	

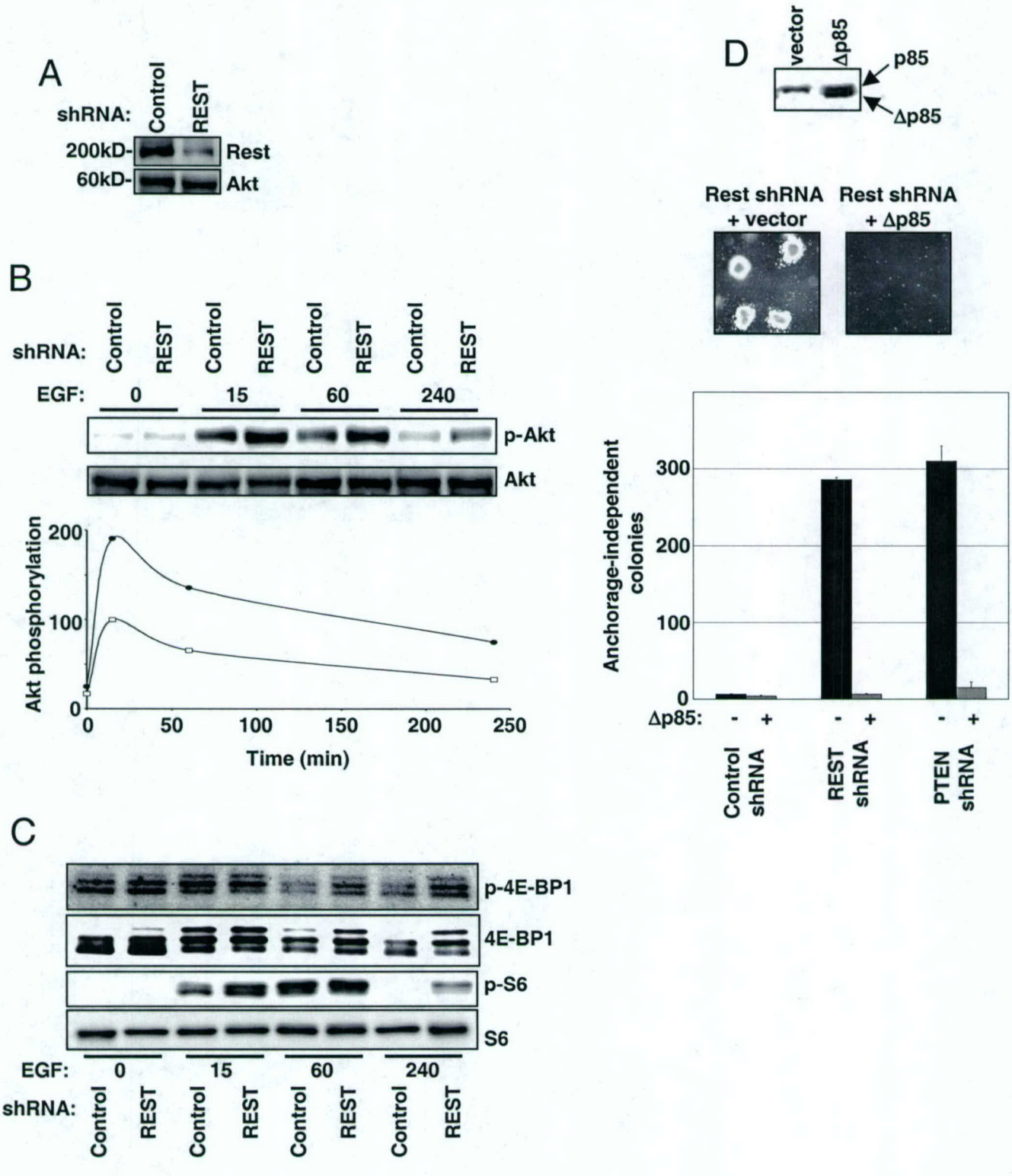
B



C




A
SW620
SW1417
200kD-
Rest
115kD-
DDB1
B
1000
100
10
Colonies
REST: - + - +
SW620 SW1417
SW1417
vector: Control Rest
C
1
2
3
4
WT
M K K S T K K K T L K N
ATG AAA AAA AGT ACA AAG AAG AAA ACT CTG AAA AAT
Colon
Mutant
ATG AAA AAA GTA CAA AGA AGA AAA CTC TGA AAA AT
M K K V Q R R K L STOP
DLD1
DBD
CBD
CoRest
D
T84
SW948
NCI H747
HCT116
DLD1
115kD-
αDDB1
85kD-
αRest-N
HMEC
REST-FS
pcDNA3
DLD-1
85kD-
αRest-N
E
200kD-
85kD-
αFlag
300
200
100
0
Anchorage-independent colonies
Control
Rest-FS
Rest


A
shRNA:
Control
REST
200kD-
Rest
60kD-
Akt
B
shRNA:
Control
REST
Control
REST
Control
REST
Control
REST
EGF:
0
15
60
240
p-Akt
Akt
Akt phosphorylation
200
100
0
0
50
100
150
200
250
Time (min)
C
p-4E-BP1
4E-BP1
p-S6
S6
EGF:
0
15
60
240
shRNA:
Control
REST
Control
REST
Control
REST
Control
REST
D
vector
Δp85
p85
Δp85
Rest shRNA
+ vector
Rest shRNA
+ Δp85
Anchorage-independent
colonies
300
200
100
0
Δp85:
-
+
-
+
-
+
Control
shRNA
REST
shRNA
PTEN
shRNA

Supplemental Table 1

Gene Name	532 Mean
Homo sapiens zinc finger protein 134 clone pHZ-15 ZNF134 mRNA.	52905
Homo sapiens dual specificity phosphatase 8 DUSP8 mRNA.	45638
Homo sapiens secretory pathway component Sec31B-1 SEC31B-1 mRNA.	35429
Homo sapiens voltage-dependent anion channel 2 pseudogene VDAC2P	25164
Homo sapiens RE1-silencing transcription factor REST mRNA.	24643
Homo sapiens spectrin beta erythrocytic includes spherocytosis	21512
Homo sapiens KIAA1361 protein KIAA1361 mRNA.	19428
Homo sapiens LIM domain only 4 LMO4 mRNA.	18552
Homo sapiens v-akt murine thymoma viral oncogene homolog 2 AKT2	15663
Homo sapiens catenin cadherin-associated protein delta 2 neural	10552
Homo sapiens stromal antigen 3 STAG3 mRNA.	9210
Homo sapiens dUTP pyrophosphatase DUT mRNA.	7395
Homo sapiens tyrosine 3-monooxygenase tryptophan 5-monooxygenase	4475
Homo sapiens cadherin 6 type 2 K-cadherin fetal kidney CDH6	3380
Homo sapiens KIAA0349 protein KIAA0349 mRNA.	2505
Homo sapiens leukocyte tyrosine kinase LTK mRNA.	2356
Homo sapiens PMS2 postmeiotic segregation increased 2 S.	2257
Homo sapiens interleukin-1 receptor-associated kinase 3 IRAK3	1979
Homo sapiens interleukin 22 receptor IL22R mRNA.	1979
Homo sapiens HMP19 protein LOC51617 mRNA.	1944
Homo sapiens fibrinogen B beta polypeptide FGB mRNA.	1891
Homo sapiens necdin homolog mouse NDN mRNA.	1795
Homo sapiens bradykinin receptor B2 BDKRB2 mRNA.	1771
Homo sapiens CGI-69 protein LOC51629 mRNA.	1734
Homo sapiens mannosidase alpha class 2A member 2 MAN2A2 mRNA.	1642
Homo sapiens coagulation factor II thrombin receptor-like 1	1625
Homo sapiens B-cell CLL lymphoma 9 BCL9 mRNA.	1541
Homo sapiens eukaryotic translation initiation factor 2B subunit 3	1433
Homo sapiens ribonuclease P 30kD RPP30 mRNA.	1387
Homo sapiens thyroid autoantigen 70kD Ku antigen G22P1 mRNA.	1307
Homo sapiens similar to Ephrin type-A receptor 5 precursor	1180
Homo sapiens myosin heavy polypeptide 11 smooth muscle MYH11	1163
Homo sapiens hypothetical protein DKFZp586I021 DKFZp586I021	952
Homo sapiens activating transcription factor 6 ATF6 mRNA.	827
Homo sapiens T brachyury homolog mouse T mRNA.	734
Homo sapiens mab-21-like 2 C. elegans MAB21L2 mRNA.	726
Homo sapiens jumping translocation breakpoint JTB mRNA.	580
Homo sapiens likely ortholog of mouse rabphilin 3A RPH3A mRNA.	494
Homo sapiens chromosome 20 open reading frame 18 C20orf18	425
Homo sapiens kinesin heavy chain member 2 KIF2 mRNA.	390
Homo sapiens LIM homeobox protein 6 LHX6 mRNA.	287

Supplemental Figure 1